

MEASURING PROCESSING SPEED DEFICITS IN MULTIPLE SCLEROSIS:
A COMPARISON OF REACTION TIME AND RAPID SERIAL PROCESSING

BY

Abbey Jean Hughes

Submitted to the Department of Psychology and the Graduate Faculty of the University of
Kansas in partial fulfillment of the requirements for the degree of Master of Arts.

Committee Members

Douglas R. Denney, Ph.D. (Chair)

David K. Johnson, Ph.D.

Sarah D. Pressman, Ph.D.

Date defended: August 18, 2010

The Thesis Committee for Abbey Jean Hughes certifies that this is the approved version of the following thesis:

MEASURING PROCESSING SPEED DEFICITS IN MULTIPLE SCLEROSIS:
A COMPARISON OF REACTION TIME AND RAPID SERIAL PROCESSING

Douglas R. Denney, Ph.D. (Chair)

Date Approved: August 18, 2010

Abstract

Research has suggested that information processing speed is the primary cognitive deficit associated with multiple sclerosis (MS). The present study featured a comparison of three paper-based and computer-based neuropsychological tests designed to measure processing speed. We found that the Simple Reaction Time subtest of the Computerized Test of Information Processing, and the combined scores for the word reading and color naming subtests of the computerized Stroop, were the most effective measures for differentiating MS patients from healthy controls in terms of processing speed. These measures also demonstrated the least susceptibility to practice effects, and the least reliance on possibly confounding cognitive processes (e.g., memory). Findings from this research will contribute to a more comprehensive understanding of the cognitive processes affected by MS, and will justify the use of computerized versions of these tests in future research.

Acknowledgements

First and foremost, I would like to thank my research advisor and thesis chair, Dr. Douglas R. Denney, not only for his dedication to this project, but also for the outstanding mentorship that he has provided during the formative years of my graduate career. I am grateful for his support throughout every stage of this work. I would also like to thank our collaborator, Dr. Sharon Lynch, for providing her invaluable clinic resources. Dr. Lynch's dedication to the treatment of people with multiple sclerosis has inspired and encouraged my research and clinical interests in this population. I would also like to thank my committee members, Dr. David Johnson and Dr. Sarah Pressman, for their helpful suggestions and insight throughout this project.

Finally, I am whole-heartedly grateful for my wonderful colleagues, friends, and family who have contributed endless support through the trials and tribulations of graduate school. I would like to emphasize my eternal appreciation and gratitude for my parents for their unfailing dedication to my development and their continued support of my education. I am fortunate to have such wonderful people in my life with whom to share this accomplishment.

Table of Contents

Title Page	1
Acceptance Page	2
Abstract	3
Acknowledgements	4
Table of Contents	5
List of Tables and Figures	6
Introduction	7
Method	17
Participants	17
Pre-screening Evaluations	17
Cognitive Measures	20
Procedure	25
Results	26
Preliminary Comparisons and Statistical Considerations	26
CTIP	28
Stroop	33
SDMT	39
Correlations	42
Discussion	47
Subtest Component Features	47
Cognitive Variability	49
Administration Format	50
Practice Effects	51
Reaction Time Versus Rapid Serial Processing	52
Relation to Disability	53
Theoretical and Clinical Implications	54
References	56

List of Tables and Figures

Tables

1. Demographic and self-report characteristics of patient and controls groups	28
2. CTIP: Comparison of MS patients and healthy controls	30
3. RSP: Comparison of MS patients and healthy controls	34
4. Comparison of MS patients and healthy controls on errors	38
5. Correlations between computer-based and paper-based RSP measures	43
6. Correlations between RT and RSP measures	44
7. Correlations between EDSS functional system scores and cognitive measures for patients	46

Figures

1. Reaction time (s) on the CTIP subtests for patients and controls	31
2. Group differences in the number of correct responses for related and unrelated words	33
3. Subinterval responses for groups on W, C, CW, and RI measures of the Stroop	37
4. Responses for patients and controls on each subinterval of the SDMT	41

Measuring Processing Speed Deficits in Multiple Sclerosis: A Comparison of Reaction Time and Rapid Serial Processing

Multiple sclerosis (MS) is a chronic, often debilitating disease affecting the central nervous system. In healthy adults, the majority of axons in the central nervous systems are insulated with myelin, a protective covering which helps electrical signals travel quickly along the axons. In MS, myelin becomes damaged through a process called demyelination, resulting in lesions that, in most cases, are distributed throughout the white matter (axons) of the central nervous system. As MS progresses, the pervasive increase in lesion number and size results in axonal destruction, causing slowed or blocked neural conductivity and communication (Charcot, 1877; Swirsky-Sacchetti, Mitchell, Seward, et al., 1992; Mammi et al., 1996). This slowing of neural communication can cause physical symptoms like muscle weakness and numbness in individuals with MS. In addition to physical symptoms, the damage can also lead to extensive cognitive problems and deficits in neuropsychological function (e.g., Zakzanis, 2000; Feinstein, 1999; Arnett et al., 1997; Wishart & Shape, 1997; Rao, St. Aubin-Faubert, & Leo, 1989; Rao, Leo, Bernardin, & Unverzagt, 1991; Hoffmann Tittgemeyer, & von Cramon, 2007).

Given that MS primarily attacks the white matter of the brain and spinal cord, it is important to note the wide range of cognitive deficits that can occur as a result of demyelination. Neural circuits within the white matter are responsible for linking cerebral cortex to spatially- and functionally-diverse regions of the brain, including areas involved in intellect, attention, memory, verbal fluency, visuo-spatial and visuo-motor activity, dexterity, sensation and perception, locomotion, affective behavior, executive function, and, speed of information processing (Kujala, Portin, Revonsuo, & Ruutiainen, 1994, 1995; Lezak, 1995; Rao, 1986; Rao et al., 1991; Ruchkin et al., 1994; Thorton & Raz, 1997; Arnett et al., 1997; Brassington &

Marsh, 1998; Zakzanis, 2000; DeLuca, Barbieri-Berger, & Johnson, 1994; Grigsby, Kaye, & Busenbark, 1994; Rao, et al., 1993; Coolidge, Middleton, Griego, & Schmidt, 1996; Beatty, Goodkin, Monson, Beatty, & Hersgard, 1988; Feinstein, 1999; Beatty & Scott, 1993). Given that people with MS suffer a devastating and insidious array of debilitation, research over the past decade has focused on characterizing these specific cognitive deficits, as well as identifying more global effects of demyelination.

As a result of demyelination, one of the most direct and widely-reported consequences of slowed neural conductivity is decreased speed of information processing (e.g., Brassington & Marsh, 1998; DeLuca, Chelune, Tulskey, Lengenfelder, & Chiaravalloti, 2004; Denney, Lynch, Parmenter, & Horne, 2004; Hoffmann Tittgemeyer, & von Cramon, 2007; Kalmar, Bryant, Tulskey, & DeLuca, 2004; Kail, 1998; Lengenfelder, Bryant, Diamond, Kalmar, Moor, & DeLuca, 2006; Schulz, Kopp, Kunkel, & Faiss, 2006). In fact, a substantial amount of research has suggested that the slowing of cognitive processing in people with MS is, quite possibly the primary deficit associated with MS (e.g., Archibald & Fisk, 2000; Bergendal, Fredrikson, & Almkvist, 2007; Bodling, Denney, & Lynch, 2008; de Sonneville et al., 2002; Demaree, DeLuca, Gaudino, & Diamond, 1999; DeLuca et al., 2004; Denney, Lynch, & Parmenter, 2008; Denney et al., 2004; Denney, Sworowski, & Lynch, 2005; Grigsby, Ayarbe, Kravcisin, & Busenbark, 1994; Kail, 1998; Kalmar et al., 2004; Lengenfelder et al., 2006; Rao et al., 1989; Rao et al., 1991; Reicker, Tombaugh, Walker, & Freedman, 2007). These studies have illustrated that, while performance is severely diminished on a wide variety of cognitive tests in MS patients, most of the variation in these tests can be specifically attributed to decreased speed of information processing. Even early in the course of the disease, MS patients exhibit significant impairments in processing speed compared to controls (Archibald & Fisk, 2000; Achiron et al., 2005;

Bergendal et al, 2007; DeLuca et al, 2004; Grigsby et al, 1994; Santiago, Guardia, Casado, Carmona, & Arbizu, 2007; Schulz et al., 2006). However, when these impairments are decoupled from other cognitive measures, differences between patients and controls significantly diminish, or disappear (Bodling et al., 2008; Denney et al., 2005; Jennekens-Schinkel, Lanser, Van Der Velde, & Sanders, 1990; Van Dijk et al., 1992; Denney & Lynch, 2009; Denney et al., 2004; Macniven et al., 2008; Vitkovitch et al., 2002). Given this prevailing emphasis on information processing speed, work in our laboratory has generally focused on characterizing, measuring, and understanding this ubiquitous cognitive deficit.

Deficits in processing speed are commonly measured using various combinations of the following neuropsychological tests: the Computerized Tests of Information Processing (CTIP; Tombaugh & Rees, 1999; Hartman, 2008; Reicker et al., 2007), the Paced Auditory Serial Addition Test (PASAT, (Gronwall, 1977), the Stroop Test (Jennekens-Schinkel et al., 1990; Rao et al., 1991; Van den Burg, van Zomeren, Minderhoud, Prange, & Meijer, 1987), and the Symbol Digit Modalities Test (SDMT; Beatty, Goodkin, Monson, & Beatty, 1989; Parmenter, Weinstock-Guttman, Garg, Munschauer, & Benedict, 2007). We will first discuss the CTIP, which is composed of three subtests designed to measure reaction time (RT). We will then discuss the latter three tests, which are designed to measure rapid serial processing (RSP). Traditionally, RT tests measure the speed at which an individual can respond to and make decisions about various types of stimuli. The CTIP is comprised of three subtests: Simple RT, Choice RT, and Semantic RT. Each subtest requires a progressive increase in the complexity of information processing. Simple RT measures an individual's baseline reaction time to a single stimulus. Choice RT requires the individual to make a discrete choice between two stimuli, which are held constant throughout the subtest. Finally, the Semantic RT subtest requires the

individual to make a conceptual or lexical choice between two stimuli, which vary throughout the subtest. The general hypothesis behind the CTIP is that as the RT subtests become more complex, they become more sensitive to slower processing speeds, such as those exhibited by people with MS (Reicker et al., 2007).

In contrast to RT tests, items in RSP tests “appear sequentially with little or no variation in the operation to be performed on each item. The operation itself is typically not very demanding, but must be executed quickly, the goal usually being to correctly complete as many items as possible in an allotted period of time” (Bodling, Denney, & Lynch, 2008). Although these tests are purported to measure factors above and beyond processing speed (e.g., focused attention, short term memory, and interference factors), research has shown that controlling for baseline processing speed diminishes or eliminates these other potential differences between MS patients and controls (e.g., Bodling et al., 2008; Denney & Lynch, 2009; Denney et al., 2004; Macniven et al., 2008; Vitkovitch et al., 2002). Importantly, these findings suggest that performance on RSP tests relies most heavily on processing speed over and above other possibly contributing factors. Recent work in our laboratory has shown differences between MS patients and controls on two of these tests, the PASAT and the Stroop Test. The PASAT is one of the more commonly used measures of RSP and is often referred to as the most sensitive measure of information processing speed for MS patients (DeLuca, Johnson., & Natelson, 1993; Tombaugh, 2006). The PASAT requires test-takers to listen to a series of single digit numbers and add each new digit to the preceding one, verbally declaring the sum before the next digit is presented. The Stroop Test involves three subtests: a word reading subtest (W), a color naming subtest (C), and a color-word naming subtest (CW). It is important to note here that work in our laboratory has used a computerized version of the Stroop (Denney & Lynch, 2009; Denney et al., 2008; Denney

et al., 2004; Denney et al. 2005). Although the stimuli presented in our version are identical to those of the paper-based version, critics have raised some concerns regarding the equivalence of these testing formats. These criticisms will be addressed later in this section. Another RSP test, the SDMT (Beatty et al., 1989; Huijbregts et al., 2004), involves the presentation of a series of nondescript geometric symbols that have been previously paired with single digit numbers (1 – 9). Upon presentation of the symbol, the test-taker must declare (orally) the digit which corresponds to the given symbol, responding to as many symbols as possible within a given timeframe. The general hypothesis behind RSP tests is that the cognitive burden of rapid, sequential responses compounds the already-diminished processing speed in MS patients. In other words, as more responses are required, RSP tests become more sensitive to slower processing speeds, such as those exhibited in people with MS.

Although both RT and RSP testing formats have been shown to successfully differentiate MS patients from controls on measures of information processing speed, there is some debate over which provides a more sensitive tool in assessing the severity of cognitive decline. Recent work in our laboratory has shown that measures requiring RSP yield larger effect sizes than measures of RT (Reicker et al. 2007), suggesting that the compounded nature of RSP tests makes them more sensitive to the slowing of cognitive processing in patients with MS (Bodling et al., 2008; Denney et al, 2008).

In addition to debate over which type of test (RT vs. RSP) yields more sensitive measures of information processing speed, there is also considerable debate over which RSP tests are most sensitive to processing speed. Past research has cited the PASAT as the “gold standard” for measuring processing speed; however, recent research has illustrated several reasons why other tests, such as the Stroop Test might be more appropriate. First, Fisk and Archibald (2001) noted

that the PASAT can be completed using a “skipping strategy,” which would allow MS patients to perform in the normal range of scores, despite having significantly lower baseline processing speeds than controls. Other criticisms of the PASAT include the propensity for subjects to exhibit practice effects across multiple administrations and for performance to be confounded with other variables such as age and education. Perhaps the most salient criticism of the PASAT for MS patients is that many subjects have reported the test to be distressing, making them less willing to complete the study (Holdwick & Wingenfled, 1999; Lezak, 2004; McCaffrey et al., 1995; Parmenter, Shucard, Benedict, & Shucard, 2006; Tombaugh, 2006). Given these potential problems with reliability, validity, and patient acceptance, it is important to consider alternative RSP measures that may present fewer methodological and theoretical obstacles.

Research in our laboratory has further challenged the PASAT’s “golden” reputation by showing that the computerized Stroop Test may be a more sensitive test of processing speed, with fewer methodological complications (e.g., Denney & Lynch, 2009). In fact, when compared to the PASAT, the computerized Stroop more effectively differentiated MS patients from controls and yielded larger effect sizes (Lynch, Dickerson, & Denney, 2010). We hypothesize that scores on the PASAT are affected not only by speed of processing, but also by non-speed related cognitive processes such as working memory, numerical skill, and focused attention. These confounding cognitive processes may contribute to the smaller effect sizes we observed compared to the Stroop. Given the Stroop’s greater efficacy at differentiating MS patients from controls, we suggest that the Stroop may be a better fit for the “gold standard” test of speed of processing. However, this assertion has not come without criticism. Traditionally, performance on the Stroop was thought to rely on two general factors: speed of processing in the initial word reading (W) and color naming (C) subtests, and resistance to interference in the third (CW)

subtest (Jenson, 1965). Under the premise that both factors significantly contribute to performance on the Stroop, critics have suggested that the effects of processing speed are diluted by the interference factor when participants are responding to the Stroop stimuli comprising the third subtest (Denney & Lynch, 2009). Indeed, we have established that MS patients perform more poorly than controls on all three subtests of the Stroop. However, the differences in performance on the CW subtest no longer remained significant when results were covaried with performance on W and/or C subtests (Denney & Lynch, 2009). Despite the original premise of the Stroop, these results suggest that processing speed is the most relevant cognitive domain distinguishing MS patients from controls.

To date, we have shown that the computerized Stroop can be a more valuable tool than the PASAT for assessing cognitive slowing in MS patients. However, we have yet to examine another RSP measure, the SDMT, and assess its efficacy in this realm. The SDMT has been used in Rao's Brief Repeatable Battery (BRB), a widely used assessment tool for cognitive evaluation of MS patients in clinical settings (e.g., Peyser, Rao, LaRocca, & Kaplan, 1990; Rao et al., 1991; Rao & Cognitive Function Study Group of the National Multiple Sclerosis Society, 1990). Unfortunately, the BRB is composed of a time-consuming repertoire of several neuropsychological tests that can be exhausting to fatigue-prone MS patients. To address this confounding issue, Portaccio et al. (2009) analyzed the relative importance and sensitivity of each test within the BRB, determined which tests were essential for detecting cognitive decline, and developed a shortened version of the BRB comprised of those essential tests. They concluded that the SDMT is indeed an essential test and should remain in the battery as an important tool for assessing cognitive deficits in MS patients. Two recent studies (Drake, Weinstock-Guttman, Morrow, Hojnacki, Munschauer, & Benedict, 2010; Forn, Belenguer,

Parcet-Ibars, & Avila, 2008) have compared the efficacy of the PASAT and the SDMT in measuring processing speed and have found the SDMT to be more sensitive to processing speed than the PASAT. In agreement with our aforementioned hypothesis, Forn et al. (2008) suggested that this may be due to the PASAT's dual emphasis on working memory and processing speed diluting its ability to purely measure the latter. Although the SDMT may be an effective measure of processing speed, conflicting studies have claimed that the SDMT may rely on other cognitive domains (e.g., working memory and attention), and may be vulnerable to practice effects (Hinton-Bayre & Geffen, 2005). More research is needed to address these criticisms, and determine whether the SDMT provides a reliable measure of processing speed.

Given that studies specifically comparing the Stroop Test to the SDMT are virtually absent from the literature, the primary aim of the present study was to assess the relative efficacy of these two tests to differentiate speed of processing in MS patients from controls. Although research has identified the SDMT as a purer measure of processing speed than the PASAT (e.g., Drake et al., 2010; Forn et al., 2008), the SDMT may still require various cognitive operations in addition to RSP, and may be sensitive to practice effects. Therefore, we are interested in the relative significance of information processing speed and other potentially confounding cognitive domains that contribute to the test's ability to distinguish MS patients from controls. Identifying the cognitive components unique to and shared by the Stroop and SDMT will allow us to determine the more sensitive measure of processing speed.

It is also important to address some of the recent criticisms regarding our laboratory's use of a computerized version of the Stroop Test. In the paper-based version of the Stroop, all stimuli in a given subtest are presented together in rows and columns on a sheet of paper. The subject is asked to scan from left to right, starting with the top row of stimuli. Conversely, the

computerized version presents each stimulus by itself in the center of the computer screen. The subject is asked to respond to the stimulus, and then press a button to advance to the next stimulus. Salo, Henik, & Robertson (2001) referred to the paper-based format as a “cluttered field,” arguing that stimuli adjacent to the test stimulus could distract the test-taker, resulting in lower performance scores on the paper-based format relative to the computerized format. Slower performance on the paper-based format may be further compounded in MS patients, who often experience visual symptoms such as optic neuritis, diplopia, nystagmus, and ophthalmoplegia. Each of these symptoms can impact ocular neuromuscular coordination in MS patients, resulting in impaired visual scanning skills. Because individual stimulus presentation does not require visual scanning, Salo et al. (2001) contended that computerized testing diminishes the effect size between clinical samples and controls. Therefore, it may be valuable to evaluate the differences in Stroop performance between MS patients and controls on both the computer-based and paper-based versions. Like the Stroop, the SDMT is traditionally administered as a paper-based, “cluttered field” test, suggesting that Salo’s criticisms may also apply to the SDMT.

Consequently, it is important to consider differences in performance between paper-based and computer-based versions of the SDMT. Given the confounding impact of physical disability on performance on paper-based tests, we contend that our computerized Stroop, and our newly-developed computerized SDMT, offer advantages over those of the paper-based versions of these tests. First, as cognitive researchers, we are primarily interested in the cognitive deficits that accompany MS. Although a paper-based test may yield larger effect sizes between MS patients and controls, these differences would predictably covary with visual impairments. We suspect that controlling for visual symptoms will eliminate their contribution to effect sizes, yielding a more accurate measure of processing speed.

In addition to providing a more refined measure of information processing speed, computerized testing also provides a more convenient, portable, and user-friendly testing format for MS patients with limited physical mobility. Moreover, the technological capabilities afforded by computers allow for superior data collection methods. For example, in the paper-based Stroop, subjects are asked to complete as many responses as possible in one 60-second trial. The experimenter times the trial and counts the number of responses. During the computer-based Stroop, the computer can not only time the entire subtest and count the number of responses, but also measure the time it takes for the subject to complete *each* response, and measure the number of responses completed in *any* subinterval of interest (e.g., every 5 seconds, the first 20 seconds, the last 10 seconds, etc). An experimenter with a stopwatch would be hard-pressed to match the precision of timing achieved by a computer.

Given the recent criticisms and benefits of our computer-based Stroop, the second aim of the present study was to investigate the equivalence between computerized and paper-based formats of the Stroop and SDMT. We hypothesized that our computerized tests will achieve sufficient equivalence. Establishing equivalence will address previous criticisms and allow for the experimentally- and clinically-relevant benefits previously discussed.

In summary, the present study had two primary objectives. First, we featured a comparison between the SDMT and the Stroop Test. The purpose was to determine the relative efficacy of these measures at distinguishing MS patients from controls on measures of processing speed. Because these tests require RSP, we also attempted to replicate earlier findings that RSP measures are more sensitive than RT measures (e.g., CTIP) to processing speed deficits in MS patients. Our second objective addressed earlier criticisms of the equivalence of the standard paper-based versions of these tests to our computerized versions of the Stroop and

SDMT. Findings from this research will help us to better understand the cognitive processes involved in the Stroop Test and SDMT, and to justify the use of our computerized versions of these tests in future research.

Method

Participants

The sample consisted of 50 patients with clinically definite MS and 40 healthy controls. Participants were between the ages of 22 and 64, with a mean of 43.76 years ($SD = 11.41$). All patients had a diagnosis of relapsing-remitting or secondary-progressive MS of at least 1 year duration and were under the care of the same neurologist (Sharon G. Lynch) at the University of Kansas Medical Center. Expanded Disability Status Scale (EDSS) scores were obtained for 42 of the patients and ranged from 0 to 8.5 with a mean of 3.57 ($SD = 2.05$). Disease durations ranged from 1 to 38 years with a mean of 10.07 years ($SD = 8.04$). Recruitment of MS patients was restricted based on the presence of any of the following conditions: neurological disorder other than MS; history of drug or alcohol abuse, psychiatric disorder, mental retardation, or traumatic head injury; current use of narcotics or benzodiazepines; visual impairment or color-blindness; relapse of MS symptoms within the past 30 days; or intellectual impairment that would interfere with comprehension of cognitive testing instructions. Recruitment of healthy controls was restricted based the following exclusion criteria: any chronic medical condition; the use of prescription medication, with the exceptions of vitamin/mineral supplements, birth control, and low-dose aspirin; and the criteria previously outlined for MS patients.

Pre-screening Evaluations

Following initial assessment and prior to cognitive testing, patients and controls completed a series of pre-screening evaluations designed to provide basic demographic and

background information. First, participants filled out a brief information sheet to determine the presence of any aforementioned exclusion criteria. Participants then completed the North American Adult Reading Test (AmNART; Grober & Sliwinski, 1991; Nelson, 1982), followed by the Fatigue Severity Scale (FSS; Krupp, LaRocca, Muir-Nash, & Steinberg, 1989) and the Center for Epidemiologic Studies – Depression Scale (CES-D; Radloff, 1977), and selected questions from the Perceived Stress Scale – 10 (PSS-10; Cohen, Kamarck, & Mermelstein, 1983). As described previously, MS patients were additionally assessed with the Expanded Disability Status Scale (EDSS; Kurtzke, 1983) in order to determine their level of disability.

Brief information sheet. Each participant completed a brief information sheet, which included his/her gender, date of birth, date first diagnosed with MS (for patients only), education level, and the presence of any exclusion criteria. Education level was rated on a 0 to 4 scale (0 = completed high school, 1 = some college, 2 = completed four-year degree, 3 = some graduate school, 4 = completed advanced graduate degree).

North American Adult Reading Test (AmNART). The AmNART, the American English version of the NART (Nelson, 1982), is one of the most common tests for assessing premorbid verbal IQ in cases of degenerative diseases such as MS, Alzheimer’s disease, traumatic brain injury, and dementia (Grober & Sliwinski, 1991). This test requires the individual to read aloud 45 irregularly spelled words and is based on the premise that word reading skills do not change dramatically in the presence of at least the mild to moderate dementia accompanying MS. The AmNART has been shown to be a reliable test in comparisons between clinical and nonclinical populations (Friend & Grattan, 1998). Ensuring equivalence between patients’ and controls’ AmNART scores allowed us to attribute differences on the cognitive testing results of the present study to the specific abilities underlying these tests rather

than to differences between the samples in general mental ability.

Fatigue Severity Scale (FSS). The FSS is a widely used questionnaire for evaluating fatigue in MS patients and was designed to differentiate between fatigue and clinical depression (Krupp et al., 1989). The questionnaire consisted of 9 short items pertaining to the severity of fatigue symptoms, with each item rated on a scale from 1 (strongly disagree) to 7 (strongly agree).

Center for Epidemiologic Studies - Depression Scale (CES-D). The CES-D is a common self-report scale for determining an individual's level of depression (Radloff, 1977). Participants completed 20 short items concerning the frequency with which they experienced depression symptoms during the preceding week. Each item was rated on a 4-point scale ranging from 1 (rarely or none of the time) to 4 (most of the time). Three questions pertaining to participants' feelings of perceived stress were included with this version of the CES-D, but were analyzed separate from CES-D questions.

Perceived Stress Scale – 10 (PSS-10). The PSS-10 is a 10-item self-report questionnaire for determining an individual's level of perceived stress. For the purposes of this study, participants completed 3 of these 10 items concerning the frequency with which they experienced stressful situations during the preceding month. Each item was rated on 4-point scale ranging from 1 (rarely or none of the time) to 4 (most of the time). Although these items were presented on CES-D questionnaire, responses were analyzed separately.

Expanded Disability Status Scale (EDSS). The EDSS is a widely used method for quantifying disability in MS patients. The EDSS assesses overall disability based on eight functional domains: pyramidal, brainstem, sensory, bowel and bladder, cerebral, and other. Assessment of each functional system yields a subscore, which is then used to calculate an

overall disability score. Overall disability scores range from 0 to 10, with 0 designating a normal neurological exam and 10 designating death due to MS.

Cognitive Measures

All participants completed a series of reaction time (RT) tests that, together, comprised the Computerized Test of Information Processing (CTIP). Participants also completed a series of rapid serial processing (RSP) tests, which consisted of a paper-based Stroop Test, a paper-based Symbol Digit Modalities Test (SDMT; one of two versions C or D), a computerized Stroop Test, and a computerized SDMT (one of two versions C or D).

Computerized Test of Information Processing (CTIP).

Simple RT (SRT). The purpose of this portion of the CTIP was to measure a participant's baseline RT. In this test, participants were instructed to respond as quickly as possible to a stimulus presented on the computer screen. The participant first received a warning stimulus (three circles), followed by the target stimulus (+) that replaces the warning stimulus. Participants were instructed to press the space bar as soon as possible when the plus sign appeared. The computer program randomly varied the time interval between the warning and target stimuli, using five randomly ordered interstimulus intervals (1000 ms, 2000 ms, 2500 ms, 3000 ms, and 4000 ms). The test began with 5 practice trials, followed by 15 test trials. The computer recorded individual RTs for the 15 test trials and computed the mean RT for these trials.

Choice RT (CRT-2). In this subtest, participants were instructed to respond as quickly as possible to one of two stimuli presented on the screen. The participant was first presented with two warning stimuli (two circles) appearing on the left and right side of the screen. After an interstimulus interval, a target stimulus (+) appeared in place of one of the

warning stimuli. Participants were instructed to press either the “F” key to designate a stimulus that appeared in place of the left circle, or the “J” key to designate a stimulus that appeared in place of the right circle. The computer program randomly varied the interstimulus interval, as described in the SRT. The test was comprised of 5 practice trials, followed by 15 test trials. The computer recorded individual RTs for each of the 15 test trials as well as whether the response to the trial was correct or incorrect. Mean RTs for all trials and for correct trials were also computed.

Semantic Search RT (SemRT). In this subtest, participants were asked to decide whether or not a word belongs to a specific semantic category. A category name (e.g., “FURNITURE”) first appeared in the middle of the screen, printed in black capital letters. After an interstimulus interval of varying length, two stimulus words appeared below the category name. One of the stimulus words constituted a member of the category name (e.g., “CHAIR”) and the other did not (e.g., “BANANA”). Participants were instructed to press “F” if the word on the left fit into the given category, or “J” if the word on the right fit into the given category. Additionally, the incorrect stimulus word (i.e., the word that did not constitute a member of the category name) was either semantically “unrelated” or “related” to the correct stimulus word. For example, a category name (e.g., “TOOL”) appeared, followed by two stimulus words, one correct (e.g., “AXE”), and one incorrect. An “unrelated” incorrect stimulus word (e.g., “BOAT”) had no association with the correct stimulus word. Conversely, a “related” incorrect stimulus word (e.g., WOOD) was associated with the correct stimulus word. In this example, “BOAT” was not related to “AXE;” however, “WOOD” was related to “AXE”. The computer randomly varied the interstimulus interval, as described in the SRT. This test was comprised of 5 practice trials, followed by 30 test trials. Of the 30 test trials, 15 contained “unrelated” stimulus words,

and 15 contained “related” stimulus words. The computer recorded individual RTs for each of the 30 test trials as well as whether the response to the trial was correct or incorrect, and whether the stimulus items were “unrelated” or “related”. Mean RTs for all trials and for correct trials were also computed.

Paper-based Stroop Test. The paper-based version of the original Stroop Test (Stroop, 1935) consisted of three white cards, each containing 20 rows of 5 items. The first card will contain randomized color names (e.g., RED, GREEN, BLUE, YELLOW) printed in black capital letters. The second card will contained a series of X’s printed in the same four colors (e.g., XXXX printed in yellow ink). The final card contained randomized color names in randomized (incongruent) colors (e.g., RED printed in blue letters). For each card, subjects were instructed to give a verbal response to the stimulus (i.e., read the word or named the color), reading each item from left to right, beginning with the first row and moving down one row at a time from top to bottom.

Prior to each subtest of the Stroop, subjects were given the following instructions: “Work quickly, but try not to make any errors. If you do make an error, try not to bother to correct your mistake. Just go on to the new item.” Each subtest lasted 60 seconds and began with an 8-item practice session followed by the full 60-second trial. The experimenter recorded the total number of responses completed and errors made for each subtest. In addition to the individual scores for the word reading (W), color naming (C), and color-word naming (CW) subtests, the experimenter calculated a combined score for word reading and color naming subtests (W+C), and a relative interference score $((C - CW)/C)$. The combined score reflected the total number of items completed for the first two subtests of the Stroop, and has been shown to be a reliable measure of processing speed (e. g., Denney & Lynch, 2009). The interference score reflected the

cognitive interference produced by the incongruity between the word and the color of the printing for the stimulus items in the CW subtest, relative to the subjects' initial performance on C subtest.

Computerized Stroop Test. A computerized version of the original Stroop Test was administered using a laptop with 14-inch screen. Like the paper-based format, the computerized Stroop consisted of three 60-second subtests, starting with the W, then C, and finally CW subtest. However, in the computerized format, each stimulus appeared individually in the center of the computer screen. Subjects were instructed to give a verbal response to each stimulus and proceed to the next stimulus by immediately pressing the space bar on the laptop. Again, subjects were instructed to respond as quickly as possible while avoiding errors. They were also told not to correct any errors that they make, but simply to proceed to the next item. Each subtest began with a practice set of 8 items, followed by the full 60-second trial. The computer timed the subtest and record the total number of responses completed in each subtest. Taking advantage of extra capabilities afforded by computerized testing, the computer also divided each 60-second subtest time into three 20-second subintervals and recorded the number of items completed in each subinterval. Combined scores (W+C) and relative interference scores were calculated according to the same formula used in the paper-based version of the Stroop Test.

Paper-based SDMT. Subjects completed one of two alternate forms (Form C or Form D) of a paper-based version of the SDMT (Smith, 1982; Hinton-Bayre et al., 1997). Each form contained a reference key, located at the top of the form, which consisted of 9 geometric shapes (symbols) specifically paired with Arabic digits 1 to 9. Below the key, the remaining portion of the form consisted of 8 rows of 15 randomized, unpaired symbols. Subjects were instructed to examine the key and give a verbal response to each unpaired stimulus (i.e., indicate the missing

digit that corresponded to the symbol). Subjects received the following instructions: “Look at the boxes at the top of the page. Each box has a symbol in it. Now look at the boxes in the row beneath the symbols. Each of those boxes has a number. Now look at the next line of boxes. Notice that each of these boxes contain a symbol, but the boxes beneath them are empty. You are to substitute that empty box with the number that should go there according to the way they are paired in the key at the top of the page and tell me what that number is. Work quickly, but try not to make any errors. If you make an error, tell me what you think the correct answer is. When you are satisfied with your answer, move on to the next item.” Each trial lasted 90 seconds and began with a 10-item practice session followed by the full 90-second trial. The experimenter recorded both the total number of items and the total number of correct items completed during the trial.

Computerized SDMT. A computerized version of the SDMT was administered using a laptop with 14-inch screen. Like the paper-based format, the computerized SDMT consisted of one 90-second trial using one of two alternate forms (Form C or Form D). Importantly, subjects who had completed Form C during the paper-based administration received Form D for the computerized administration, and vice versa, in order to eliminate practice effects. However, in contrast to the paper-based format, each stimulus in the computerized format appeared individually in the center of the computer screen. The key remained visible at the top of the screen throughout the trial. Subjects were instructed to use the key to pair the stimulus with its corresponding digit, give a verbal response to each stimulus, and proceed to the next stimulus by immediately pressing the space bar on the laptop. Again, subjects were instructed to respond as quickly as possible while avoiding errors. They were also told that they can correct any error they happened to make before moving on to the next item. Testing began with a practice set of 10 items, followed by the full 90-second test. The computer timed the trial and recorded the total

number of responses, while the experimenter recorded the total number of correct responses. As an additional advantage to computerized testing, the computer also divided each 90-second test time into three 30-second subintervals and record the number of items completed in each subinterval.

Procedure

Before any recruitment began, this study was approved by the Human Subjects Committees at the University of Kansas Medical Center and the University of Kansas-Lawrence. Eligible MS patients were initially introduced to the study during the course of their regular appointments in the MS Clinic. The patient first met with the research assistant who introduced the study. If the patient expressed an initial willingness to participate, the research assistant obtained written consent. The patient then met with Dr. Sharon Lynch who assessed his/her disability with the EDSS and proceeded with their regular exam. If time permitted after the exam, the research assistant then proceeded with the remaining prescreening evaluations and cognitive testing. If, however, patients preferred to complete the cognitive testing session at a later date, the research assistant scheduled an appointment for testing either at the MS Clinic or in their homes.

Controls were recruited from the metropolitan areas of Lawrence, Kansas City, KS, Kansas City, MO, and St. Louis, MO. Recruitment occurred from among personal acquaintances of the researchers and the patients (e.g., their spouses and adult children) and other contacts with the community (e.g., personal appeals to church groups and other service organizations for volunteers to participate in the study). Interested parties, if eligible, were instructed to contact the graduate research assistant to schedule a cognitive testing session either at the MS Clinic or in the participant's home.

Each testing session lasted approximately 40-45 minutes, and consisted of a series of pre-screening evaluations, followed by a series of cognitive tests. Each participant completed all four RSP tests (the standard paper-based Stroop Test, the standard paper-based SDMT, the computerized version of the Stroop Test, and the computerized SDMT), with test order and format counterbalanced across individuals. Because the SDMT may be prone to practice effects, we used alternate forms of the SDMT (Forms C and D), and counterbalanced forms across format conditions (paper vs. computer). The three RT tests (SRT, CRT, and SemRT) were administered between paper-based and computer-based administrations of RSP tests. Cognitive testing was followed by a short debriefing. Participants' questions regarding the study were addressed during the debriefing and participants received contact information for any future questions that might arise.

Results

All 40 healthy adults completed the present study. Of the 50 patients who consented to participate, eight participants dropped out of the study between consent procedures and cognitive testing, yielding a final sample of 42 patients who completed the study. Demographic information was collected from the eight patients who did not complete cognitive testing and analyses were conducted to ensure equivalence between completers and non-completers. There were no significant differences between patients who dropped out and those who continued (all $ps > .05$). The most commonly reported reason for attrition was participants' having insufficient time to complete the study and this occurred predominantly for participants who had scheduled appointments in their homes.

Preliminary Comparisons and Statistical Considerations

Demographic and self-report characteristics of the patients and control groups are shown

in Table 1. Patients and controls did not differ with respect to gender, age, perceived stress (PSS-10), or scores on the AmNART. However, with respect to education, patients reported lower levels of education than controls. Given these differences, all analyses of cognitive measures were adjusted by including education level as a covariate. Education level did not emerge as significant covariate for any cognitive test included in this study. Therefore, this variable was not included in the final analyses.

Patients and controls also differed with respect to self-reported scores on fatigue and depression (Table 1), with patients having higher scores than controls on the FSS and CES-D. Differences in fatigue and depression scores required consideration when comparing groups' performance on cognitive measures. For all analyses, FSS and CES-D scores were first entered as covariates. For the majority of analyses described below, neither measure emerged as a significant covariate. Moreover, because elevated fatigue and depression are inherent features of MS and not merely the results of random sampling error during study recruitment, subsequent analyses exclude the contribution of these factors to the differences in cognitive performance.

Table 1
Demographic and self-report characteristics of patient and controls groups

	MS Patients (N = 42)		Controls (N = 40)		<i>t</i> (df = 80)	<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Gender (F/M)	33/9		29/11		0.41	.52
Age	44.83	12.09	42.63	10.70	0.87	.38
Education Level	1.79	1.34	2.45	1.50	2.12	.04
FSS	37.88	15.12	17.28	9.23	7.49	< .001
CES-D	34.36	10.47	26.30	5.08	4.47	< .001
PSS-10	6.81	2.27	6.38	2.36	0.85	.40
AmNART	30.36	8.01	32.20	5.81	1.20	.24

* Chi-square analysis

** Based on Fisher's Exact Test

CTIP

Raw RT scores for each reaction time test were examined for outliers. Extremely fast responses, defined as less than 150 ms, were removed from the data because they likely reflected anticipatory responding. Extremely slow responses were also calculated, eliminated, and replaced according to the following procedure: 1) for each subtest of the CTIP, group means and standard deviations were calculated across all RT scores and all participants in that group; 2) any individual RT that exceeded its group mean by three or more standard deviations was deleted; 3) a regression procedure was used to replace each missing value with a predicted value that was interpolated on the basis of linear regression.

Table 2 illustrates the means and *SD* for patients and controls on each score derived from the CTIP. Obtained *t* values, *p* values, and effect sizes (Cohen's *d*) based on simple independent *t* tests comparing the means for the patients versus the controls on each variable are also presented

in Table 2. Responses on both the CRT and the SemRT could be correct or incorrect. In order to determine whether reaction times on correct items alone were different than those on all items, a 2 (Group) X 2 (Accuracy) repeated measures ANOVA was performed on participants' reaction times on the CRT and likewise, on the SemRT. The main effect for accuracy was not significant for either subtest (CRT: $F(1, 80) = 2.06, p > .05, \eta^2 = .03$; SemRT: $F(1, 80) = 1.53, p > .05, \eta^2 = .02$). Likewise, there was no significant Group X Accuracy interaction for either subtest (CRT: $F(1, 80) = 0.003, p > .05, \eta^2 < .001$; SemRT: $F(1, 80) = 2.53, p > .05, \eta^2 = .03$). The correlation between reaction times for correct items and for all items was 1.00 ($p < .001$) on the CRT and .675 ($p < .001$) on the SemRT. Based on these findings, no further attention was given to possible differences in reaction times for correct items, and all analyses reported here were performed on reaction times to all items.

Table 2
CTIP: Comparison of MS patients and healthy controls

	MS Patients (N = 42)		Controls (N = 40)		<i>t</i> (df = 80)	<i>p</i>	Cohen's <i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
SRT							
RT:							
Total items	0.40	0.10	0.31	0.04	5.32	< .001	1.16
CRT							
RT:							
Total items	0.50	0.13	0.40	0.05	4.72	< .001	1.02
Correct items	0.50	0.13	0.40	0.05	4.72	< .001	1.02
Accuracy (# correct):							
Total items	14.86	0.52	14.95	0.22	1.06	.29	0.23
SemRT							
RT:							
Total items	1.23	0.33	1.00	0.14	4.44	< .001	0.96
Rel items	1.29	0.36	1.02	0.15	4.36	< .001	0.95
Unrel items	1.17	0.31	0.94	0.15	4.26	< .001	0.93
Correct items	1.22	0.33	0.98	0.14	4.39	< .001	0.95
Rel items	1.28	0.36	1.03	0.14	4.27	< .001	0.93
Unrel items	1.16	0.31	0.93	0.15	4.27	< .001	0.93
Accuracy (# correct):							
Total items	29.64	0.69	29.32	0.83	1.89	.06	0.42
Rel	14.86	0.42	14.52	0.60	2.90	.01	0.65
Unrel	14.79	0.52	14.80	0.52	0.13	.90	0.03

The subtests of the CTIP increased in complexity. In order to assess the impact of task complexity on participants' reaction times, a 2 (Group) X 3 (Complexity) repeated measures analysis of variance (ANOVA) was conducted. Analyses revealed significantly longer RTs for the patients than controls on all subtests ($F(1, 80) = 25.63, p < .001, \eta^2 = .24$) and progressively longer RTs for all participants as the subtests increased in complexity ($F(2, 160) = 904.67, p < .001, \eta^2 = .92$). Furthermore, as complexity increased across the three subtests, the patients' RTs progressively diverged from those of controls, resulting in a significant Group X Complexity interaction ($F(2, 160) = 10.37, p < .001, \eta^2 = .12$). Figure 1 illustrates the performances of patients and controls across each subtest of the CTIP.

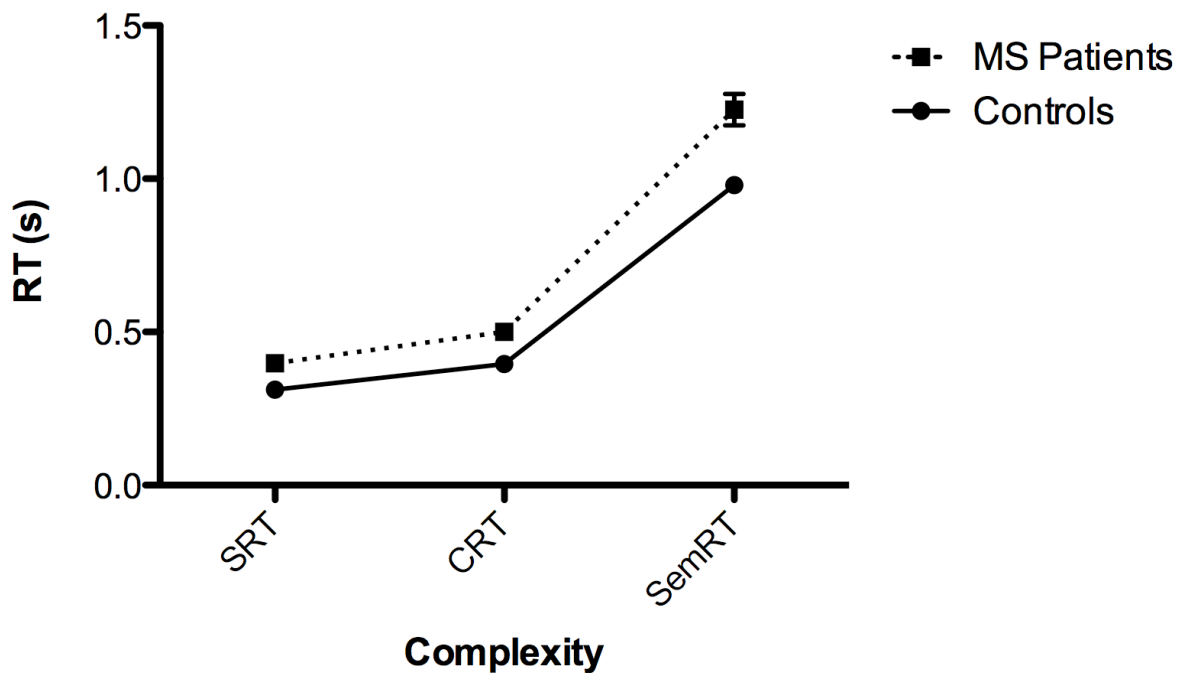


Figure 1. Reaction times (s) on the CTIP subtests for patients and controls.

The items of the SemRT were divided between those in which the incorrect stimulus word was semantically related or unrelated to the correct stimulus word. In order to determine whether semantic relatedness on the SemRT had an impact on participants' performance, 2 (Group) X 2 (Relatedness) analyses were conducted on the reaction times and the number of correct responses on this subtest. Analyses revealed significantly longer RTs for related words than for unrelated words ($F(1, 80) = 52.86, p < .001, \eta^2 = .40$); however, the number of correct responses were not significantly different for related and unrelated words ($F(1, 80) = 1.75, p > .05, \eta^2 = .02$). Moreover, although there was no significant Group X Relatedness interaction for reaction times ($F(1, 80) = 1.58, p > .05, \eta^2 = .02$), there was a significant interaction for the number of correct responses $F(1, 80) = 5.08, p < .05, \eta^2 = .06$. As illustrated in Figure 2, patients completed more correct items than controls when words were related, but there were no differences between groups when words were unrelated. Of note, this group difference was the only instance in which there was a significant group difference in accuracy of performance on the CTIP (Table 2).

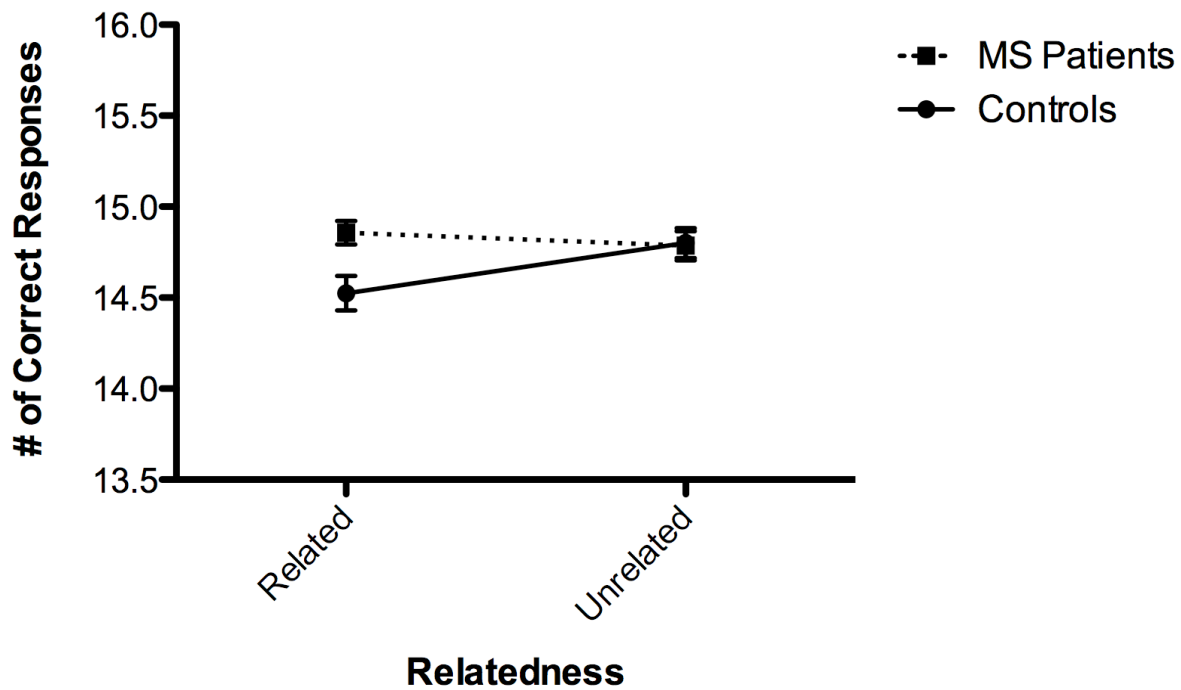


Figure 2. Group differences in the number of correct responses for related and unrelated words.

Stroop

Table 3 illustrates the means and *SD* for patients and controls on each score derived from the Stroop. Scores were based on the number of correct items for each subtest. All subsequent analyses were performed on the following five scores: W, C, CW, W+C, and RI. And, as in the case of the CTIP, obtained *t* values, *p* values, and effect sizes (Cohen's *d*) based on simple independent *t* tests comparing the means for the patients versus the controls on each variable derived from the Stroop are also presented in Table 3.

Table 3
RSP: Comparison of MS patients and healthy controls

	MS Patients (N = 42)		Controls (N = 40)		<i>t</i> (df = 80)	<i>p</i>	Cohen's <i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
Stroop Test							
Computer-based							
W	70.38	18.12	90.48	16.92	5.18	< .001	1.15
C	59.69	11.17	75.52	11.97	6.20	< .001	1.37
CW	43.71	10.89	54.12	9.81	4.54	< .001	1.00
W+C	130.07	28.17	166.00	27.64	5.83	< .001	1.29
RI	27.40	9.04	27.91	10.24	0.24	.81	0.05
Paper-based							
W	109.40	26.18	139.38	21.75	5.62	< .001	1.24
C	78.55	17.17	101.68	14.03	6.66	< .001	1.47
CW	48.90	14.24	62.42	11.75	4.68	< .001	1.03
W+C	187.95	41.32	241.05	31.33	6.53	< .001	1.44
RI	37.93	12.26	38.42	9.30	0.20	.84	0.05
SDMT							
Computer-based	51.43	10.90	66.60	9.52	6.70	< .001	1.48
Paper-based	47.02	12.34	62.90	8.46	6.76	< .001	1.49

Separate 2 (Group) X 2 (Format) ANOVAs were performed for each variable derived from the Stroop to determine if differences in performance occurred between computerized and paper-based formats. Analyses of word reading (W), color naming (C), combined scores (W+C), and color-word naming (CW) revealed significant main effects for Group (W: $F(1, 80) = 37.36$, $p < .001$, $\eta^2 = .32$; C: $F(1, 80) = 50.47$, $p < .001$, $\eta^2 = .39$; W+C: $F(1, 80) = 46.71$, $p < .001$, $\eta^2 = .37$; CW: $F(1, 80) = 24.61$, $p < .001$, $\eta^2 = .24$) and Format (W: $F(1, 80) = 391.23$, $p < .001$, $\eta^2 = .83$; C: $F(1, 80) = 285.27$, $p < .001$, $\eta^2 = .78$; W+C: $F(1, 80) = 458.80$, $p < .001$, $\eta^2 = .85$; CW: $F(1, 80) = 45.98$, $p < .001$, $\eta^2 = .37$), with patients completing fewer items than controls and all participants completing fewer items on the computer-based test than on the paper-based test.

Analyses revealed a significant Group X Format interaction for word reading, color naming, and combined scores (W: $F(1, 80) = 4.94, p < .05, \eta^2 = .06$; C: $F(1, 80) = 7.49, p < .01, \eta^2 = .09$; W+C: $F(1, 80) = 7.65, p < .01, \eta^2 = .09$), with a greater disparity between patients and controls on the paper-based than on the computer-based format. This interaction was not significant in the case of the color-word naming score ($F(1,80) = 2.44, p > .05, \eta^2 = .03$). Analysis of interference yielded a significant main effect for Format ($F(1, 80) = 87.15, p < .001, \eta^2 = .52$), with the computerized version yielding lower interference scores than the paper-based version. However, neither the main effect for Group ($F(1, 80) = 0.06, p > .05, \eta^2 = .001$) nor the Group X Format interaction ($F(1, 80) < .001, p > .01, \eta^2 < .001$) was significant.

Each participant completed the Stroop Test two times during the course of the procedure and the order in which the formats (i. e., paper-based vs. computer-based) were administered was counterbalanced across participants. Therefore, it was possible to examine participants' performance across the two administrations of the Stroop to determine whether there was evidence of practice effects across the two trials. Separate 2 (Group) X 2 (Trial) ANOVAs were applied to each Stroop variable for this purpose. Significant main effects for Group were already established on the basis of prior analyses, and given the focus on practice effects, only the main effect and interaction term involving Trial were considered in these analyses. The main effect for Trial was not significant in the case of word reading, color naming, or combined scores (W: $F(1, 80) = 0.01, p > .05, \eta^2 < .001$; C: $F(1,80) = 0.04, p > .05, \eta^2 = .001$; W+C: $F(1,80) = 0.02, p > .05, \eta^2 < .001$). However, for the color-word naming and interference scores, all participants completed significantly more items and obtained lower interference scores on the second administration (CW: $F(1,80) = 7.76, p < .01, \eta^2 = .09$; RI: $F(1,80) = 5.15, p < .03, \eta^2 = .06$), suggesting that practice effects impacted performance on this more challenging subtest. For all

five Stroop variables, Group X Trial interactions were not significant (W: $F(1, 80) = 0.05, p > .05, \eta^2 = .001$; C: $F(1,80) = 0.03, p > .05, \eta^2 < .001$; W+C: $F(1,80) = 0.01, p > .05, \eta^2 < .001$; CW: $F(1, 80) = 0.52, p > .05, \eta^2 = .01$).

Each subtest of the computer-based Stroop was divided into three 20-second subintervals and the number of correct responses for each subinterval was recorded. Therefore, it was possible to examine participants' performance across a subtest in order to determine whether there was evidence of practice effects within the trial. Separate 2 (Group) X 3 (Subinterval) ANOVAs were applied to each Stroop variable for this purpose. Figure 3 illustrates participants' performance across each subinterval of the word reading, color naming, and color-word reading subtests, as well as participants' interference scores across each subinterval. As in the case of the Group X Trial analyses, significant main effects for Group were already established on the basis of prior analyses, and, therefore, only the main effect and interaction term involving Subinterval are considered in these analyses. Analyses indicated that participants' performance did not significantly differ across the subintervals composing the word reading, color naming, or combined scores (W: $F(2, 160) = 1.21, p > .05, \eta^2 = .02$; C: $F(2,160) = 1.73, p > .05, \eta^2 = .02$; W+C: $F(2, 160) = 0.23, p > .05, \eta^2 = .003$). However, scores on color-word naming and interference, did differ significantly across subintervals (CW: $F(2, 160) = 3.71, p < .05, \eta^2 = .04$; RI: $F(2, 160) = 5.40, p < .01, \eta^2 = .06$). *Post hoc* comparisons using the Fisher's Least Significant Difference (LSD) procedure revealed that participants completed significantly fewer items and obtained greater interference scores during the first 20-second subinterval compared to the second and third subintervals (all $ps < .05$); however, there were no significant differences in performance or interference between the second and third 20-second subintervals ($ps > .05$), suggesting that acclimation to the task occurs early during this subtest and then remains steady

thereafter. For all five Stroop variables, Group X Subinterval interactions were not significant (W: $F(2, 160) = 0.20, p > .05, \eta^2 = .003$; C: $F(2, 160) = 0.10, p > .05, \eta^2 = .001$; W+C: $F(2, 160) = 0.06, p > .05, \eta^2 = .001$; CW: $F(2, 160) = 0.8, p > .05, \eta^2 = .01$).

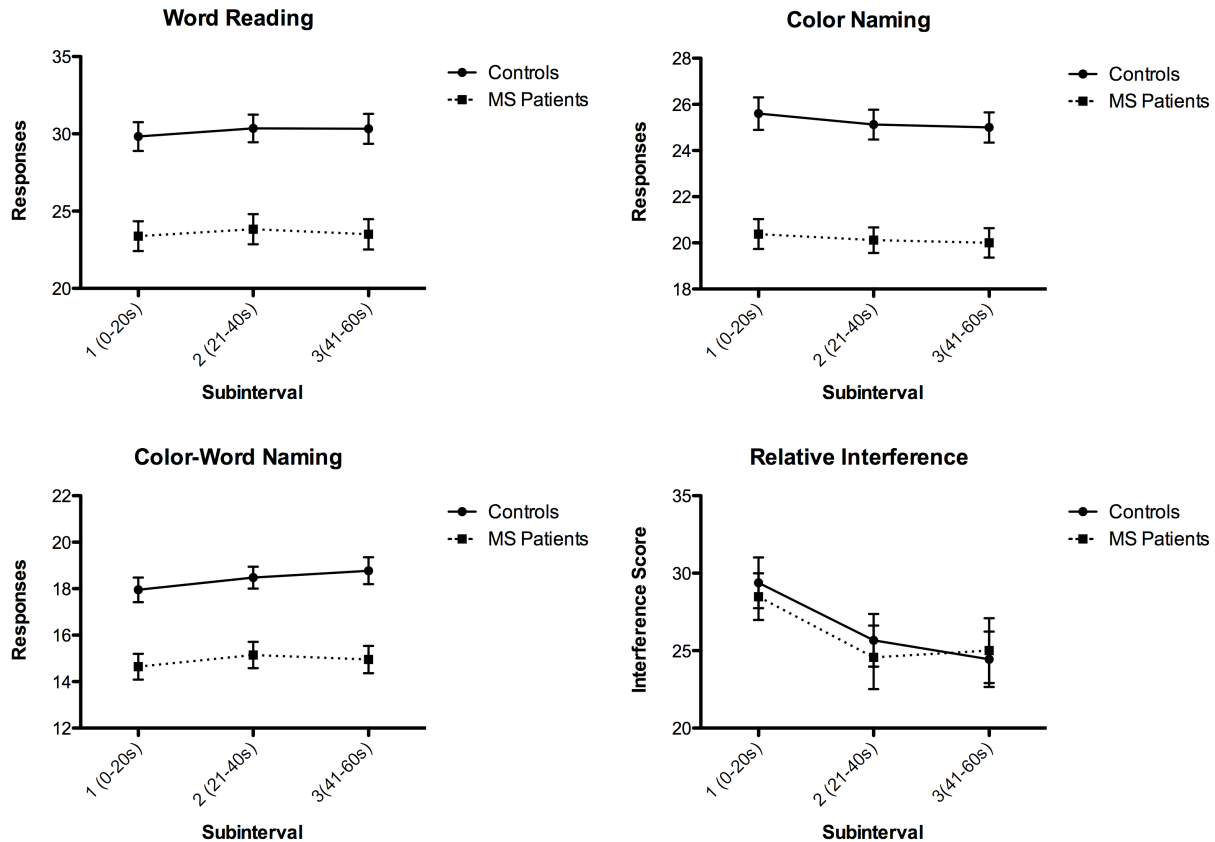


Figure 3. Subinterval responses for groups on W, C, CW, and RI measures of the Stroop.

Up to this point, all analyses for Stroop variables have been based on the total correct items each participant completed in a given trial or subinterval. These scores were computed by taking the total number of items completed in a trial and subtracting the number of errors. Previous research has not examined whether MS patients and controls differ in the number of errors made in a given trial. The following analyses examined these differences. Table 4 compares the means and *SD* for patients and controls on the number of errors for each subtest of

the Stroop and for the combined score. And, as in previous tables, p values and effect sizes (Cohen's d) based on simple independent t tests comparing the mean errors for the patients versus the controls are also presented in Table 4.

Table 4
RSP: Comparison of MS patients and healthy controls on errors

	MS Patients (N = 42)		Controls (N = 40)		t (df = 80)	p	Cohen's d
	M	SD	M	SD			
Stroop Test							
Computer-based							
W	0.33	0.79	0.02	.16	2.49	.02	0.54
C	0.81	1.19	0.20	1.19	3.07	.003	0.67
CW	1.02	1.47	1.08	1.46	0.16	.88	0.03
W+C	1.14	1.69	0.22	0.53	3.35	.002	0.72
Paper-based							
W	1.00	1.94	0.15	0.43	2.77	.01	0.60
C	0.93	1.55	0.40	0.78	1.96	.05	0.43
CW	1.62	1.86	1.15	1.44	1.27	.21	0.28
W+C	1.93	2.65	0.55	0.99	3.16	.003	0.68
SDMT							
Computer-based	0.31	0.90	0.38	1.33	0.26	.79	0.06
Paper-based	0.88	1.45	0.58	1.06	1.09	.28	0.24

Separate 2 (Group) X 2 (Format) ANOVAs were performed for errors on each subtest, as well as for the combined score. Analyses for word reading, color naming, and combined scores revealed a main effect for Group (W: $F(1, 80) = 9.95$, $p < .01$, $\eta^2 = .11$; C: $F(1, 80) = 9.41$, $p < .01$, $\eta^2 = .11$; W+C: $F(1, 80) = 15.44$, $p < .001$, $\eta^2 = .16$), with controls making fewer errors than patients on both formats. Groups did not differ significantly on the color-word naming subtest ($F(1, 80) = 0.50$, $p > .05$, $\eta^2 = .01$). Analyses for word reading and the combined scores also

revealed a significant main effect for Format (W: $F(1, 80) = 6.42, p < .05, \eta^2 = .08$; W+C: $F(1,80) = 5.85, p < .05, \eta^2 = .07$), with participants making more errors on the paper-based Stroop than on the computerized version. Format did not significantly impact the number of errors on color naming and color-word naming subtests (C: $F(1, 80) = 1.09, p > .05, \eta^2 = .01$; CW: $F(1,80) = 3.33, p > .05, \eta^2 = .04$). Finally, there were no significant Group X Format interactions (W: $F(1, 80) = 3.13, p > .05, \eta^2 = .04$; C: $F(1,80) = 0.07, p > .05, \eta^2 = .001$; W+C: $F(1, 80) = 1.01, p > .05, \eta^2 = .01$; CW: $F(1,80) = 2.01, p > .05, \eta^2 = .02$).

SDMT

Table 3 presents the means and *SD* for patients and controls on the number of correct items completed during the paper-based and computer-based versions of the SDMT. As in the case of the other cognitive tests, obtained *t* values, *p* values, and effect sizes (Cohen's *d*) based on simple independent *t* tests comparing the means for the patients versus the controls are also presented in Table 3.

A preliminary 2 (Group) X 2 (Form) repeated measures ANOVA was performed to determine whether Form C and Form D were equivalent. The focus of this analysis was on the main effect for Form and the interaction term. Differences between groups are considered in the main analyses to be reported below. Neither the main effect for Form ($F(1, 80) = 0.62, p > .05, \eta^2 = .01$), nor the Group X Form interaction ($F(1, 80) = 0.50, p > .05, \eta^2 = .01$) was significant, suggesting that forms of the SDMT were equivalent.

A 2 (Group) X 2 (Format) ANOVA was performed to determine if differences in performance occurred between computer-based and paper-based formats. Analysis revealed significant main effects for Group ($F(1, 80) = 50.05, p < .001, \eta^2 = .39$) and Format ($F(1, 80) = 32.52, p < .001, \eta^2 = .29$), with patients completing fewer items than controls and all participants

completing fewer items on the computer-based test than on the paper-based test. Analyses did not yield a significant Group X Format interaction ($F(1, 80) = 0.25, p > .05, \eta^2 = .003$).

Each participant completed the SDMT two times during the course of the procedure and the order in which the formats were administered was counterbalanced across participants. Therefore, it was possible to examine participants' performance across the two administrations of the SDMT to determine whether there was evidence of practice effects across the two trials. A 2 (Group) X 2 (Trial) ANOVA indicated that performance by all participants did not change over the course of the two administrations of the SDMT ($F(1, 80) = 0.06, p > .05, \eta^2 = .001$). Furthermore, the Group X Trial interaction was not significant ($F(1, 80) = 0.18, p > .05, \eta^2 = .002$).

The computerized SDMT was divided into three 30-second subintervals and the number of correct responses for each subinterval was recorded. Therefore, it was possible to examine participants' performance across the 90-second trial in order to determine whether there was evidence of practice effects within the trial. Figure 4 illustrates participants' performance across the three subintervals that made up the SDMT. Significant main effects for Group were already established on the basis of prior analyses, and given the focus on practice effects, only the main effect and interaction term involving Subinterval were considered in this analysis. A 2 (Group) X 3 (Subinterval) ANOVA revealed that performance significantly differed across subintervals ($F(1, 80) = 24.44, p < .001, \eta^2 = .23$). *Post hoc* analyses using the Fisher's LSD procedure demonstrated that participants completed progressively more items during each successive 30-second subinterval (all $ps < .005$). This progressive improvement in performance for all participants over the 90-second test was qualified by a significant Group X Subinterval interaction ($F(2, 160) = 3.45, p < .05, \eta^2 = .04$), indicating that groups differed in the rate at

which this improvement occurred. Analyses revealed that controls' performance increased predominantly between the first and second 30-second subinterval. In contrast, patients' performance increased predominantly between the second and third 30-second subinterval, suggesting a slower acquisition of the symbol-digit associations for the patients compared to controls.

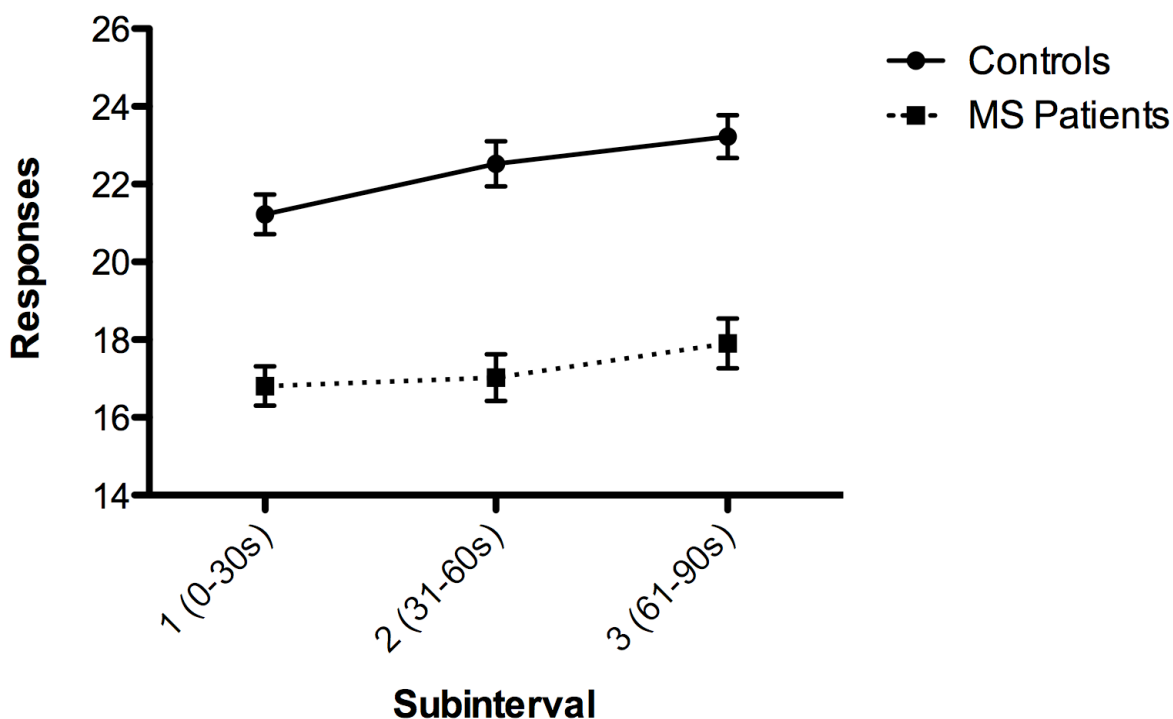


Figure 4. Responses for patients and controls on each subinterval of the SDMT.

Prior to this point, all analyses for the SDMT have been based on the total correct items each participant completed in a given trial or subinterval. These scores were computed by taking the total number of items completed in a trial and subtracting the number of errors. As in the case with the Stroop, the present section will examine whether MS patients and controls differed in the number of errors made in a given trial. To further explore the influence of forms (i.e. Form C

and Form D) and formats (i.e. computerized and paper-based), main effects and interaction terms for these factors are considered in the analysis. Table 4 compares the means and *SD* for patients and controls on the number of errors for each format of the SDMT. And, as in previous tables, *p* values and effect sizes (Cohen's *d*) based on simple independent *t* tests comparing the mean errors for the patients versus the controls are also presented in Table 4.

Each participant completed the SDMT two times during the course of the procedure and the order in which the forms were administered was counterbalanced across testing formats. Therefore it was possible to examine participants' performance across the two formats of the SDMT to determine whether form version influenced the number of errors made. A 2 (Group) X 2 (Form) ANOVA revealed no significant differences between groups ($F(1, 80) = 1.01, p > .05, \eta^2 = .01$); however there was a main effect of Form ($F(1, 80) = 13.41, p < .001, \eta^2 = .14$), with participants making significantly more errors on Form D than Form C. There was no significant Group X Format interaction ($F(1, 80) = 1.71, p > .05, \eta^2 = .02$).

A 2 (Group) X 2 (Format) ANOVA was conducted to determine if format influenced the number of errors made on each administration of the SDMT. Given that significant main effects for Group were already established on the basis of the Group X Form analyses, only the main effect and interaction term involving Format are considered in these analyses. Neither the main effect for Format ($F(1, 80) = 3.89, p > .05, \eta^2 = .04$), nor the Group X Format interaction ($F(1, 80) = 0.90, p > .05, \eta^2 = .01$) was significant, suggesting that administration format did not influence the number of errors made on a given trial.

Correlations

In order to assess correlations between cognitive measures for patients and controls, bivariate correlations were computed separately for each group and then transformed

into z scores using Fisher's z' transformation procedure for combining correlations across the groups. For the purpose of the present study, the focus of these analyses was on the relationship between computerized and paper-based RSP formats, and the relationship between RT and RSP measures.

Pearson Product Moment Correlation Coefficients (R) were calculated and combined for participants' performance on the following computerized and paper-based measures: total W, C, CW, W+C, and RI scores derived from the Stroop; and total scores derived from the SDMT. The resulting correlations between computerized and paper-based measures ranged from .048 to .739. Pearson's R and p values for these correlations are presented in Table 5.

Table 5
Correlations between computer-based and paper-based RSP measures

	Computer-based	
	R	p
Paper-based		
Stroop Test		
W	.569	< .001
C	.659	< .001
CW	.726	< .001
W+C	.657	< .001
RI	.522	< .001
SDMT	.818	< .001

Coefficients were calculated and combined for participants' performance on the following RT and RSP measures: total SRT, total CRT, total SemRT, related SemRT, and unrelated SemRT scores on the CTIP; W, C, CW, W+C, RI scores on the computerized and paper-based Stroop; and total scores on the computerized and paper-based SDMT. The resulting correlations between RT and computerized RSP measures ranged from .071 to .456, and those between RT and paper-based RSP measures ranged from .043 to .465. Pearson's R and p values for these correlations are presented in Table 6.

Table 6
Correlations between RT and RSP measures

	RT: CTIP									
	SRT		CRT		SemRT		RelRT		UnrelRT	
	R	p	R	p	R	p	R	p	R	p
RSP (Computer-based)										
Stroop										
W	-.372	.001	-.324	.003	-.355	.001	-.318	.004	-.369	.001
C	-.412	< .001	-.383	< .001	-.452	< .001	-.419	< .001	-.456	< .001
CW	-.326	.003	-.433	< .001	-.393	< .001	-.368	.001	-.397	< .001
W+C	-.405	< .001	-.364	.001	-.407	< .001	-.371	.001	-.416	< .001
RI	.071	.526	.296	.007	.178	.110	.181	.104	.169	.128
SDMT	-.401	< .001	-.351	.001	-.355	.001	-.303	.006	-.383	< .001
RSP (Paper-based)										
Stroop										
W	-.416	< .001	-.407	< .001	-.448	< .001	-.421	< .001	-.445	< .001
C	-.457	< .001	-.450	< .001	-.345	.002	-.318	.004	-.355	.001
CW	-.263	.017	-.398	< .001	-.276	.012	-.225	.042	-.312	.004
W+C	-.465	< .001	-.455	< .001	-.436	< .001	-.408	< .001	-.439	< .001
RI	-.101	.366	.129	.248	.073	.515	.043	.704	.097	.386
SDMT	.216	.052	-.404	< .001	-.282	.010	-.215	.052	-.334	.002

In order to assess the relationship between patients' disability status and their performance on cognitive measures, a series of bivariate correlations were performed on the patient group alone. Spearman Rank Order Correlation Coefficients (R_s) were calculated for disability and cognitive variables. Disability variables consisted of patients' overall EDSS scores, six functional system scores, and disease duration, measured in years. Cognitive variables consisted of patients' performance on total items completed on the SRT, CRT, SemRT subtests of the CTIP; total W, C, CW, W+C, and RI scores derived from the computerized and paper-based Stroop; and total items completed on the computerized and paper-based SDMT. The resulting correlations between disability and CTIP scores ranged from .013 to .542, those between disability and computerized RSP measures ranged from .007 to .627, and those between disability and paper-based RSP measures ranged from .027 to .723. Spearman's R_s and p values for these correlations are presented in Table 7.

Table 7
Correlations between EDSS functional system scores and cognitive measures for patients

EDSS		Visual		Nystagmus		Sensory		Brainstem		Pyramidal		Cerebellar		Duration	
<i>R_s</i>	<i>p</i>	<i>R_s</i>	<i>p</i>	<i>R_s</i>	<i>p</i>	<i>R_s</i>	<i>p</i>	<i>R_s</i>	<i>p</i>	<i>R_s</i>	<i>p</i>	<i>R_s</i>	<i>p</i>	<i>R_s</i>	<i>p</i>
RT Measures															
CTIP															
SRT	.522 < .001	.046	.774	.244	.120	.277	.075	.320	.039	.467	.002	.313	.044	.031	.845
CRT	.601 < .001	.078	.622	.176	.264	.293	.059	.331	.032	.550 < .001		.357	.020	.013	.936
SemRT	.410 .007	.053	.739	.284	.068	.235	.133	.542 < .001		.310 .046		.302	.052	-.049	.759
RSP Measures (Computer-based)															
STROOP															
W	-.544 < .001	.069	.665	-.084	.596	-.417	.006	-.266	.088	-.554 < .001		-.478	.001	-.192	.222
C	-.460 .002	.056	.726	-.130	.411	-.299	.055	-.326	.035	-.434 .004		-.340	.028	.049	.759
CW	-.447 .003	.028	.862	-.141	.373	-.221	.159	-.340	.028	-.384 .012		-.392	.010	-.007	.964
W+C	-.530 < .001	.048	.764	-.101	.523	-.387	.011	-.300	.054	-.514 < .001		-.456	.002	-.117	.460
RI	.313 .044	-.004	.982	.217	.168	.009	.953	.331	.032	.239 .127		.368	.016	.094	.553
SDMT	-.568 < .001	.085	.592	-.271	.082	-.274	.079	-.453	.003	-.505 .001		-.627 < .001		-.229	.145
RSP Measures (Paper-based)															
STROOP															
W	-.492 .001	.040	.800	-.122	.441	-.420	.006	-.361	.019	-.476 .001		-.416	.006	-.041	.797
C	-.489 .001	.111	.485	-.176	.265	-.326	.035	-.276	.077	-.467 .002		-.406	.008	.027	.868
CW	-.464 .002	.153	.335	-.389	.011	-.238	.130	-.459	.002	-.422 .005		-.391	.011	-.132	.405
W+C	-.508 .001	.084	.598	-.173	.272	-.375	.015	-.348	.024	-.489 .001		-.419	.006	.027	.866
RI	.164 .301	-.142	.369	.326	.035	-.036	.823	.312	.044	.133 .402		.167	.290	.283	.069
SDMT	-.723 < .001	.082	.607	-.224	.154	-.429	.005	-.379	.013	-.650 < .001		-.629 < .001		-.418	.006

Discussion

Previous studies using a variety of cognitive measures have consistently offered converging evidence for a dramatic decline in processing speed as the primary cognitive deficit associated with MS. The major findings from the present study further characterized and expanded on this conclusion. Significant differences between groups on the CTIP, Stroop, and SDMT suggested that all three cognitive measures effectively distinguished patients from controls in terms of processing speed. However, analyses yielded several important differences between these tests with respect to subtest component features (i.e., efficacy versus complexity), cognitive variability, administration format (i.e., computer versus paper), susceptibility to practice effects, task type (i.e. RT versus RSP), and relation to disability.

Subtest Component Features

Before considering overall differences between the three cognitive measures, we must examine the relative strengths and weaknesses of each subtest of these measures in order to determine if a particular subtest yielded stronger associations with the grouping variable than the test as a whole. For the CTIP, results suggested that all three subtests, the SRT, CRT, and SemRT, distinguished patients from controls. However, when considering the subtests separately, several differences emerged, and not all CTIP subtests were equally effective at distinguishing patients from controls in terms of processing speed.

For all three subtests, analyses of variance demonstrated that group membership had a significant impact on reaction time. However, in terms of more practical or clinical utility, the SRT yielded the greatest effect size difference between groups. These results replicated earlier findings (Reicker et al., 2007), suggesting that the SRT may be more effective than the CRT and SemRT subtests in terms of sensitivity. Across groups, an analysis of variance demonstrated that

subtest complexity had a significant impact on reaction time, and that this relationship was qualified by a significant interaction between group and subtest complexity. However, although the main effect for complexity was substantiated by a large effect size, the effect size for the interaction was considerably smaller. Because the impact of test complexity was more robust than the impact of group membership or the interaction between these two factors, we believe it to be likely that the CRT and SemRT subtests require participants to recruit brain regions over and above those required for speeded information processing alone. For example, in addition to information processing, both the CRT and SemRT subtests require resistance to interfering stimuli (e.g., incorrect choice responses), and, the SemRT subtest also requires a search of semantic memory. We argue that these added cognitive demands essentially dilute participants' scores on the CRT and SemRT as pure measures of processing speed, and, therefore, we cannot determine whether the larger group differences observed for these more complex subtests are due to processing speed, resistance to interference, semantic memory, language, or an interaction of various other cognitive factors. Consistent with Reicker et al. (2007), these results suggest that, in addition to sensitivity, the SRT may be more effective than the CRT and SemRT in terms of what might be termed "construct specificity." Consequently, further discussion of RT measures will be restricted to the SRT subtest of the CTIP.

For the Stroop, significant differences between patients and controls occurred for W, C, CW, and W+C scores. However, as in the case of the CTIP, differences in effect sizes from these analyses warranted further consideration. The present study replicated earlier findings (e. g., Denney & Lynch, 2009), showing that differences in RI scores were not significant, and effect size differences for CW scores were smaller than those for W, C, or W+C scores. These results suggested that at least one additional cognitive domain, complex selective attention, impacts CW

scores beyond that of mere processing speed. We argue that this complex selective attention, like the more complex cognitive processes involved in the CRT and SemRT subtests of the CTIP, essentially dilutes the CW scores as a pure measure of processing speed. Given that relative interference scores provide a fairly pure measure of complex selective attention independent of processing speed, and that groups exhibited no differences in this additional cognitive domain, the effect size differences for CW scores were also diluted. These results provided further support for the use of W+C scores as the primary measure of processing speed to be derived from the Stroop. Consequently, further discussion of Stroop measures will be restricted to W+C scores.

Cognitive Variability

For both the CTIP and Stroop, we have introduced the idea that, as task complexity increases, group differences in performance also increase statistically. For example, the SemRT subtest yielded greater differences between group means relative to the SRT (Figure 1). This is the common form taken by a cognitive load analysis, with an interaction between task complexity and group membership purportedly confirming the cognitive load effect. However, an examination of the effect size differences between groups revealed that those for more complex tasks were smaller than for simpler tasks. The principle reason for these seemingly contradictory findings is that effect size is affected not only by differences between group means, but also by overall variability in the data. Indeed, results demonstrated that as subtest complexity increased, so did the overall variability in the data, as illustrated by greater standard errors on the SemRT relative to the SRT (Figure 1). Considering that analysis of variance does not sufficiently capture this variability, and, moreover, that the F statistic is heavily influenced by sample size, analysis of effect size provides a clinically more meaningful measure of group differences that is

both relative to variability and independent of sample size. Given these results, simpler measures, like the SRT, are ideal for maximizing the impact of group differences in processing speed, while minimizing the impact of variability on estimates of effect size.

Administration Format

For the Stroop and SDMT, differences in participants' performance between computerized and paper-based administration formats raised important considerations for future work with these tests. Across groups, participants completed more items on the paper-based than on the computer-based versions of each test, with this difference being larger for patients than for controls on the Stroop. However, despite these statistically significant main effects and interaction, effect size differences between groups were essentially equivalent, and large, for the W+C scores and SDMT scores obtained with both formats. One should also consider the "purity" or construct sensitivity of the scores obtained under each format. It would seem that the paper-based format of the Stroop and SDMT place additional physical and cognitive demands that are not imposed by computerized formats of these tests. The paper-based formats consist of a single sheet of paper, with all items presented at one time. The participant must physically scan the entire page, moving their eyes from left to right, and from top to bottom, for the entire task. This requirement puts physical strain on the eyes, which may confound with deficits in oculomotor control frequently associated with MS. Conversely, because the computerized version administers one item in the center of the screen, there is less of a physical demand for scanning the screen. Secondly, the paper-based version presents all items together on a single page, allowing participants to "cheat" by looking ahead to the next item before they complete their response to the previous item, and it is unclear whether groups differ in their use of this anticipatory strategy. Conversely, because the computerized version only allows advancement to

the next item once the previous response has been given, “cheating” is not an option. Given the advantages discussed above, the computerized format of the W+C and SDMT scores provided the most ideal measures for maximizing sensitivity and retaining construct specificity when assessing group differences in processing speed.

Practice Effects

Practice effects for the Stroop and SDMT were assessed in two ways: test-retest comparisons and within-trial comparisons. Because they made use of the computer’s capability of segmenting the total trial time into subintervals, the within-trial comparisons were possible only for the computerized versions of the Stroop and SDMT. For the Stroop, participants did not improve or worsen between the first and second administration for W, C, and W+C scores, suggesting that there are no practice effects when these preliminary subtests are repeated on separate occasions. Additionally, participants did not improve or worsen between the three subintervals for these scores, suggesting that there are no practice effects within a single administration for these scores. However, participants did exhibit practice effects across administrations and within a single administration for CW and RI scores. Participants completed fewer CW items and had greater RI scores for the first relative to the second administration, and for the first 20-second subinterval relative to the remaining 40 seconds of work time within a single administration. These results suggested that the complex selective attention component of the CW subtest, which is not present in the preliminary subtests, may be sensitive to practice effects both between administrations and within a single administration. Given these results, we argue that the simpler subtests, particularly W+C scores, are not only more effective in terms of differentiating patients and controls, but also in avoiding confounding due to practice effects both between administrations and within a single administration. For the SDMT, participants did

not improve or worsen between the first and second administration; however participants did improve within each trial, progressively completing more items during each successive 30-second subinterval. Interestingly, controls improved more between the first and second subinterval, while patients improved more between the second and third subinterval, suggesting that there a type of “learning curve” associated with the SDMT, but that the slope of this curve may differ between patients and controls. Consequently, results of the present study suggest that while the W+C subtest of the Stroop and the SDMT are resistant to practice effects between multiple administrations, the SDMT may be prone to practice effects within a single administration.

Reaction Time Versus Rapid Serial Processing

Correlations between SRT, W+C, and SDMT scores were assessed to determine the extent to which tests that utilize measures of reaction time were related to measures of rapid serial processing. Computerized W+C and SDMT scores and paper-based W+C scores were highly correlated with SRT scores, with lower scores on these rapid serial processing measures being associated with longer RTs. Interestingly, paper-based SDMT scores were not significantly correlated with SRT scores, providing further support for some of the criticisms of paper-based formats described earlier. In addition to correlations, group differences in effect size were compared to determine which test type (i. e., RT or RSP) was more effective at differentiating patients from controls. In congruence with previous research, results of the present study suggested that RSP measures, particularly W+C and SDMT scores, yielded greater effect sizes than RT measures, particularly SRT scores. However, given the confounding variables associated with the SDMT, it is likely that W+C scores provided the purest and most effective measure of processing speed, compared to other RSP measures and to RT measures.

Relation to Disability

Consistent with previous research, all cognitive measures examined in the present study were significantly correlated with total EDSS ratings of disability, with the exception of the RI scores from the Stroop. Correlations were also examined for functional systems and disease duration. We were particularly interested in assessing the relation between cognitive performance and visual impairment. In terms of visual impairment, we were most interested in visual acuity, which was measured on an ordinal scale from the Visual functional system, and the presence of nystagmus, which was measured on a dichotomous scale from the Brainstem functional system. Previous research in our laboratory has suggested that performance on the Stroop is not affected by the levels of visual acuity and nystagmus manifested by the patients with MS that populate our studies. The present study replicated these findings and extended them to performance on the SRT and SDMT.

MS patients typically exhibit more visual impairment than healthy adults, and this deficit may impact performance on visually demanding tasks. In the present study, the oculo-motor demands of the paper-based formats were considerably greater than those of the computerized formats. As a result, we were particularly interested in how patients' visual acuity and nystagmus ratings might correlate with performance on the paper-based measures. We found no such associations in the present study. Consistent with previous studies in our laboratory, these results are likely due to patients exhibiting very little variability in visual acuity and nystagmus ratings, thereby attenuating any possible correlation. This explanation is also supported by the fact that the effect size differences between groups were essentially equivalent for both paper-based and computerized formats.

EDSS pyramidal system scores were highly correlated with SRT, W+C and SDMT scores. Additionally, W+C and SDMT scores were highly correlated with EDSS cerebellar system scores. Correlation with disability can be viewed as both a positive and negative feature of a given cognitive measure. As Lynch, Denney, & Dickerson (2010) noted, MS has pervasive physical and cognitive consequences for which it is reasonable to expect concomitant decline in both domains of functioning. However, if decline in these separate domains are too highly correlated, there exists some redundancy in measuring both and it becomes difficult to separate physical from cognitive deficits. Therefore, it is important to consider if all physical measures are correlated to all cognitive measures, or if there is enough separation to warrant separate assessments. As we have shown in recent studies (Bodling et al., 2008), and in the present study, not all measures of physical disability were highly correlated with cognitive performance, particularly with regard to visual function and disease duration. We can therefore conclude that SRT, W+C, and SDMT scores make a sufficiently unique contribution to the overall evaluation of disability status.

Theoretical and Clinical Implications

In conclusion, results from the present study suggest that processing speed differences are most sensitively and selectively measured with scores from the SRT, W+C, and, although more arguably, the SDMT. W+C and SRT scores may be most clinically useful in terms of detecting cognitive deficits early in the disease process, before other clinical domains are impacted. However, because these measures are designed to impact such a targeted, and perhaps, narrow, cognitive domain, it is very likely necessary to assess other forms of cognitive impairment. Other investigators have emphasized the impact of MS on working memory, verbal processing, and executive function; however, we contend that dysfunction in these domains cannot be assessed

without first extricating the impact of processing speed. Therefore, assessments are needed to target these domains without involving processing speed, or more likely, utilizing separate measures of processing speed that may be used to adjust scores on these other tests. The latter approach demands the purest measures of processing speed one can identify, and the present study suggests these might lie with the simpler forms of reaction time trials and with the preliminary trials of the Stroop.

References

- Achiron, A., Polliack, M., Rao, S. M., Barak, Y., Lavie, M., Appelboim, N., et al. (2005). Cognitive patterns and progression in multiple sclerosis: Construction and validation of percentile curves. *Journal of Neurology, Neurosurgery and Psychiatry*, 76, 744-749.
- Archibald, C. J., & Fisk, J. D. (2000). Information processing efficiency in patients with multiple sclerosis. *Journal of Clinical and Experimental Neuropsychology*, 22, 686-701.
- Arnett, P. A., Rao, S. M., Grafman, J., Bernardin, L., Luchetta, T., Binder, J. R., & Lobeck, L. (1997). Executive functions in multiple sclerosis: an analysis of temporal ordering, semantic encoding, and planning abilities. *Neuropsychology*, 11 (4), 535-544.
- Beatty, W. W., Goodkin, D. E., Monson, E., Beatty, P. A., & Hersgard, D. (1988). Anterograde and retrograde amnesia in patients with chronic-progressive multiple sclerosis. *Archives of Neurology*, 45, 611-619.
- Beatty, W. W., Goodkin, D. E., Monson, N., Beatty, P. A. (1989). Cognitive disturbances in patients with relapsing remitting multiple sclerosis. *Archives of Neurology*, 46, 1113-1119.
- Beatty, W. W., & Scott, J. G. (1993). Issues and developments in the neuropsychological assessment of patients with multiple sclerosis. *Journal of Neurologic Rehabilitation*, 7, 87-97.
- Bergendal, G. Fredrikson, S., & Almkvist, O. (2007). Selective decline in information processing in subgroups of multiple sclerosis: An 8-year longitudinal study. *European Neurology*, 57, 193-202.
- Bodling, A. M., Denney, D. R., & Lynch, S. G. (2008). Rapid serial processing in patients with multiple sclerosis: The contribution of peripheral deficits. *Journal of the International*

- Neuropsychological Society*, 14, 646-650.
- Brassington, J. C. & Marsh, N. V. (1998). Neuropsychological aspects of multiple sclerosis. *Neuropsychology Review*, 8 (2), 43-77.
- Charcot, J. M. (1877). *Lectures on the diseases of the nervous system delivered at La Salpetriere (translated from French)*. London: The New Sydenham Society.
- Cohen, S., Karmarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, 24, (385-396).
- Coolidge, F. L., Middleton, P. A., Griego, J. A., & Schmidt, M. M. (1996). The effects of interference on verbal learning in multiple sclerosis. *Archives of Clinical Neuropsychology*, 11, 605-611.
- De Sonneville, L. M., Boringa, J. B., Reuling, I. E., Lazeron, R. H., Adèr, H. J., & Polman, C. H. (2002). Information processing characteristics in subtypes of multiple sclerosis. *Neuropsychologia*, 40, 1751-1765.
- DeLuca, J., Barbieri-Berger, S., & Johnson, S. K.. (1994). The nature of memory impairments in multiples sclerosis: Acquisition versus retrieval. *Journal of Clinical and Experimental Neuropsychology*, 16 (2), 183-189.
- DeLuca, J., Chelune, G. J., Tulskey, D. S., Lengenfelder, J., & Chiaravalloti, N. D. (2004). Is speed of processing or working memory the primary information processing deficit in multiple sclerosis? *Journal of Clinical and Experimental Neuropsychology*, 26 (4), 550-562.
- DeLuca, J., Johnson, S. K., & Natelson, B. H. (1993). Information processing efficiency in chronic fatigue syndrome and multiple sclerosis. *Archives of Neurology*, 50, 301-304.
- Demaree, H. A., DeLuca, J., Gaudino, E. A., & Diamond, B. J. (1999). Speed of processing as

- the key deficit in multiple sclerosis: Implications for rehabilitation. *Journal of Neurology, Neurosurgery, and Psychiatry*, 67, 661-663.
- Denney, D. R., & Lynch, S. G. (2009). The impact of multiple sclerosis on patients' performance on the Stroop Test: Processing speed vs. interference. *Journal of the International Neuropsychological Society*, 15, 451-458.
- Denney, D. R., Lynch, S. G., & Parmenter, B. A. (2008). A 3-year longitudinal study of cognitive impairment in patients with primary progressive multiple sclerosis: Speed matters. *Journal of the Neurological Sciences*, 267, 129-136.
- Denney, D. R., Lynch, S. G., Parmenter, B. A., & Horne, N. (2004). Cognitive impairment in relapsing and primary progressive multiple sclerosis: Mostly a matter of speed. *Journal of the International Neuropsychological Society*, 10, 948-956.
- Denney, D. R., Sworowski, L. A., & Lynch, S. G. (2005). Cognitive impairment in three subtypes of multiple sclerosis. *Archives of Clinical Neuropsychology*, 20, 967-981.
- Drake, A. S., Weinstock-Guttman, B., Morrow, S. A., Hojnacki, D., Munshauer, F. E., & Benedict, R. H. B. (2010). Psychometrics and normative data for the Multiple Sclerosis Functional Composite: replacing the PASAT with the Symbol Digit Modalities Test. *Multiple Sclerosis*, 16, 228-237.
- Feinstein, A. (1999). *The Clinical Neuropsychiatry of Multiple Sclerosis*. New York: Cambridge University Press.
- Fisk, J. D., & Archibald, C. J. (2001). Limitation of the Paced Auditory Serial Addition Test as a measure of working memory in patients with multiple sclerosis. *Journal of the International Neuropsychological Society*, 7, 367-372.
- Forn, C., Belenguer, A., Parcet-Ibars, M. A., & Avila, C. (2008). Information-processing speed is

- the primary deficit underlying the poor performance of multiple sclerosis patients in the Paced Auditory Serial Addition Test (PASAT). *Journal of Clinical and Experimental Neuropsychology*, 30 (7), 789-796.
- Friend, K. B. & Grattan, L. (1998). Use of the North American Adult Reading Test to Estimate Premorbid Intellectual Function in Patients with Multiple Sclerosis. *Journal of Clinical and Experimental Neuropsychology*, 20 (6), 846-851.
- Grigsby, J., Ayarbe, S. D., Kravcisin, N., & Busenbark, D. (1994). Working memory impairment among persons with chronic progressive multiple sclerosis. *Journal of Neurology*, 241, 125-131.
- Grober, E. & Sliwinski, M. (1991). Development and validation of a model for estimating premorbid verbal intelligence in the elderly. *Journal of Clinical and Experimental Neuropsychology*, 13, 933-949.
- Gronwall, D. (1977). Paced Auditory Serial-Addition Task: A measure of recovery from concussion. *Perceptual and Motor Skills*, 44, 367-373.
- Hartman, D. E. (2008). The Computerized Test of Information Processing (CTIP by Tom Tombaugh. *Applied Neuropsychology*, 15 (3), 1532-4826.
- Hinton-Bayre, A. & Geffen, G. (2005). Comarability, Reliability, and Practice Effects on Alternate Forms of the Digit Symbol Substitution and Symbol Digit Modalities Tests. *Psychological Assessment*, 17 (2), 237-241.
- Hoffmann, S., Tittgemeyer, M., & von Cramon, D. (2007). Cognitive impairment in multiple sclerosis. *Current Opinion in Neurology*, 20 (3), 275-280.
- Holdwick, D. J., & Wingenfeld, S. A. (1999). The subjective experience of PASAT testing: Does the PASAT induce negative mood? *Archives of Clinical Neuropsychology*, 14, 273-284.

- Huijbregts, S. C. J., Kalkers, N. F., de Sonnevile, L. M. J., de Groot, V., Reuling, I. E. W., & Polman, C. H. (2004). Differences in cognitive impairment of relapsing remitting, secondary, and primary progressive MS. *Neurology*, 63, 335-339.
- Jennekens-Schinkel, A., Lanser, J. B. K., Van Der Velde, E. A., & Sanders, E. A. C. M. (1990). Performances of multiple sclerosis patients in tasks requiring language and visuocognition: Assessment of outpatients in quiescent disease stages. *Journal of the Neurological Sciences*, 95, 89-103.
- Jenson, A. R. (1965). Scoring the Stroop test. *Acta psychological*, 24, 398-408.
- Kail, R. (1997). The neural noise hypothesis: Evidence from processing speed in adults with multiple sclerosis. *Aging, Neuropsychology, and Cognition*, 4, 157-165.
- Kalmar, J., Bryant, D., Tulsky, D., & DeLuca, J. (2004). Information processing deficits in multiple sclerosis: Does choice of screening instrument make a difference? *Rehabilitation Psychology*, 49 (3), 213-218.
- Krupp, L. B., LaRocca, N. G., Muir-Nash, J., & Steinberg, A. D. (1989). The fatigue severity scale: Application to patients with multiple sclerosis and systemic lupus erythematosus. *Archives of Neurology*, 46 (10), 1121-1123.
- Kujala, P., Portin, R., Revonsuo, A., Ruutinen, J. (1994). Automatic and controlled information processing in multiple sclerosis. *Brain*, 117, 1115-1126.
- Kujala, P., Portin, R., Revonsuo, A., Ruutinen, J. (1995). Attention related performance in two cognitively different subgroups of patients with multiple sclerosis. *Journal of Neurology, Neurosurgery, and Psychiatry*, 59, 77-82.
- Kurtzke, M. D. (1983). Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS). *Neurology*, 33, 1444.

- Lengenfelder, J., Bryant, D., Diamond, B., Kalmar, J., Moor, N., & DeLuca, J. (2006). Processing speed interacts with working memory efficiency in multiple sclerosis. *Archives of Clinical Neuropsychology*, 21, 229-238.
- Lezak, M. D. (1995). *Neuropsychological assessment* (3rd ed.). New York: Oxford University Press.
- Lezak, M. D. (2004). *Neuropsychological assessment* (4th ed.). New York: Oxford University Press.
- Lynch, S. G., Dickerson, K. J., & Denney, D. R. (in prep). Evaluating speed in multiple sclerosis: A comparison of two rapid serial processing measures.
- Macniven, J. A., Davis, C., Ho, M. Y., Bradshaw, C. M., Szabadi, E., & Constantinescu, C. S. (2008). Stroop performance in multiple sclerosis: information processing, selective attention, or executive functioning? *Journal of the International Neuropsychological Society*, 14, 805-814.
- Mammi, S., Filippi, M., Martinelli, V., Campi, A., Colombo, B., Scotti, G., Canal, N., & Comi, G. (1996). Correlation between brain MRI lesion volume and disability in patients with multiple sclerosis. *Acta Neurologica Scandinavica*, 94, 93-96.
- McCaffrey, R. J., Cousins, J. P., Westervelt, H. J., Martynowicz, M., Remick, S. C., Szebenyi, S., et al. (1995). Practice effects with the NIMH AID abbreviated neuropsychological battery. *Archives of Clinical Neuropsychology*, 10, 241-250.
- Nelson, H. E. (1982). *National Adult Reading Test (NART): Test manual*. Windsor, UK: NFER Nelson.
- Parmenter, B. A., Shucard, J. L., Benedict, R. H. B., & Shucard, D. W. (2006) Working memory deficits in multiple sclerosis: comparison between the n-back task and the Paced Auditory

- Serial Addition Test. *Journal of the International Neuropsychological Society*, 12, 677-687.
- Parmenter, B. A., Weinstock-Guttman, B., Garg, N., Munschauer, F., & Benedict, R. H. B. (2007). Screening for cognitive impairment in multiple sclerosis using the Symbol Digit Modalities Test. *Multiple Sclerosis*, 13, 52-57.
- Peyser, J. M., Rao, S. M., LaRocca, N. G., & Kaplan, E. (1990). Guidelines for Neuropsychological Research in Multiple Sclerosis. *Archives of Neurology*, 47 (1), 94-97.
- Portaccio, E., Goretti, B., Zipoli, V., Siracusa, G., Sorbi, S., & Arnato, M. P. (2009). A short version of Rao's Brief Repeatable Battery as a screening tool for cognitive impairment in multiple sclerosis. *The Clinical Neuropsychologist*, 23 (2), 268-275.
- Radloff, L. S. (1977). The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1 (3), 385-401.
- Rao, S. M. (1986). Neuropsychology of multiple sclerosis: a critical review. *Journal of Clinical and Experimental Neuropsychology*, 8 (5), 503-542.
- Rao, S. M. & Cognitive Function Study Group, NMSS. (1990). *A manual for the Brief Repeatable Battery of Neuropsychological Tests in Multiple Sclerosis*. New York: National Multiple Sclerosis Society.
- Rao, S. M., Grafman, J., DiGiulio, D., Mittenberg, W., Bernardin, L., Leo, G. J., Luchetta, T., & Unverzagt, F. (1993). Memory dysfunction in multiple sclerosis: its relation to working memory, semantic encoding, and implicit learning. *Neuropsychology*, 7, 364-374.
- Rao, S. M., Leo, G. J., Bernardin, L., & Unverzagt, F. (1991). Cognitive dysfunction in multiple sclerosis. I. Frequency, patterns, and prediction. *Neurology*, 41, 685-691.

- Rao, S. M., St. Aubin-Faubert, P., & Leo, G. L. (1989). Information processing speed in patients with multiple sclerosis. *Journal of Clinical and Experimental Neuropsychology*, 11 (4), 471-477.
- Reicker, L. I., Tombaugh, T. N., Walker, L., & Freedman, M. S. (2007). Reaction time: An alternative method for assessing the effects of multiple sclerosis on information processing speed. *Archives of Clinical Neuropsychology*, 22, 655-664.
- Ruchkin, D. S., Drafinan, J., Krauss, G. L., Johnson, R., Canoune, H., & Ritter, W. (1994). Event-related brain potential evidence for verbal working memory deficit in multiple sclerosis. *Brain*, 117, 289-305.
- Salo, R., Henik, A., & Robertson, L. C. (2001). Interpreting Stroop Interference: An Analysis of Differences Between Task Versions. *Neuropsychology*, 15 (4), 462-471.
- Santiago, O., Guardia, J., Casado, V., Carmona, O., & Arbizu, T. (2007). Specificity of frontal dysfunctions in relapsing-remitting multiple sclerosis. *Archives of Clinical Neuropsychology*, 22, 623-629.
- Schulz, D., Kopp, B., Kunkel, A., & Faiss, J. H. (2006). Cognition in the early stage of multiple sclerosis. *Journal of Neurology*, 253, 1002-1010.
- Smith, A. (1982). *Symbol Digit Modalities Test- Revised*. Los Angeles: Western Psychological Services.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643-662.
- Swirsky-Sacchetti, T., Mitchell, D. R., Seward, J., Gonzales, C., Lublin, F., Knobler, R., & Field, H. L. (1992). Neuropsychological and structural brain lesions in multiple sclerosis: a regional analysis. *Neurology*, 42, 1291-1295.

- Thornton, A. E., & Raz, N. (1997). Memory impairment in multiple sclerosis: a quantitative review. *Neuropsychology*, 11 (3), 357-366.
- Tombaugh, T. N. (2006). A comprehensive review of the Paced Auditory Serial Addition Test (PASAT). *Archives of Clinical Neuropsychology*, 21, 53-76.
- Tombaugh, T. N., & Rees, L. (1999). *Computerized Test of Information Processing (CTIP)*. Unpublished test. Ottawa, Ontario, Canada: Canada.
- Van den Burg, W., van Zomeren A. H., Minderhoud, J. M., Prange, A. J. A., & Meijer, N. S. A. (1987). Cognitive impairment in patients with multiple sclerosis and mild physical disability. *Archives of Neurology*, 44, 494-501.
- Van Dijk, J. G., Jennekens-Schindel, A., Caekebeke, J. F. V., & Zwinderman, A. H. (1992). Are event-related potentials in multiple sclerosis indicative of cognitive impairment? Evoked and event-related potentials, psychometric testing and response speed: a controlled study. *Journal of the Neurological Sciences*, 109, 18-24.
- Vitkovitch, M., Bishop, S., Dancey, C., & Richards, A. (2002). Stroop interference and negative priming in patients with multiple sclerosis. *Neuropsychologica*, 40, 1560-1574.
- Wishart, H. & Sharpe, D. (1997). Neuropsychological aspects of multiple sclerosis: a quantitative review. *Journal of Clinical and Experimental Neuropsychology*, 19 (6), 810-824.
- Zakzanis, K. K. (2000). Distinct neurocognitive profiles in multiple sclerosis subtypes. *Archives of Clinical Neuropsychology*, 15 (2), 115-136.